The pivotal role of causality in local quantum physics

1999 J. Phys. A: Math. Gen. 32 5937

(http://iopscience.iop.org/0305-4470/32/32/309)

J. Phys. A: Math. Gen. **32** (1999) 5937–5974. Printed in the UK PII: S0305-4470(99)03671-9

The pivotal role of causality in local quantum physics*

Bert Schroer†
Institut für Theoretische Physik, FU-Berlin, Arnimallee 14, 14195 Berlin, Germany
E-mail: schroer@cbpfsu1.cat.cbpf.br

Received 20 April 1999

Abstract. In this paper an attempt is made to present very recent conceptual and computational developments in quantum field theory (QFT) as a new manifestation of old well-established physical principles. The vehicle for converting quantum algebraic aspects of local quantum physics into more classical geometric structures is the modular theory of Tomita. As the laureate E H Wichmann, together with his collaborators showed for the first time in sufficient generality, its use in physics goes beyond Einstein causality. This line of research recently gained momentum when it was realized that it is not only of great structural and conceptual innovative power, but also promises a new computational road into nonperturbative QFT which, figuratively speaking, enters the subject on the extreme opposite (noncommutative) side relative to (Lagrangian) quantization.

1. Introduction

Among the fundamental physical principles of this century which have stood their ground in the transition from classical to quantum physics, relativistic causality as well as the closely related locality of quantum operators (together with the localization of quantum states) is certainly the most prominent one.

This principle entered physics through Einstein's 1905 special relativity, which in turn resulted from bringing the Galilean relativity principle of classical mechanics into tune with Maxwell's theory of electromagnetism. Therefore, it incorporated Faraday's 'action at a neighbourhood' principle which revolutionized 19th century physics.

The two different aspects of Einstein's special relativity, namely Poincaré covariance and the locally causal propagation of waves in Minkowski space were kept together in the classical setting. In the adaptation of relativity to LQP (local quantum physics‡), on the other hand [1], it is appropriate to keep them, at least initially, apart in the form of positive-energy representations of the Poincaré group (leading to Wigner's concept of particles) and Einstein's causality of local observables (leading to observable local fields and local generalized 'charges'). Here a synthesis is also possible, but it occurs on a deeper level than in the classical setting and results in LQP as a new physical realm which is conceptually very different from both classical field theory and general QT (quantum theory). The elaboration of this last point constitutes one of the aims of this paper. We pay particular attention to those aspects of LQP which are not within the reach of standard quantum physical intuition.

* Dedicated to Professor Eyvind H Wichmann on the occasion of his 70th birthday.
† Present address: CBPF, Rua Dr Xavier Sigaud, 22290-180 Rio de Janeiro, Brazil.
‡ We use this terminology whenever we want to make clear that we relate the principles of QFT with a different formalism than that based on quantization through Lagrangian formalism.

The most remarkable aspect of quantum field theory (QFT) in its more than 60-year existence, in addition to its great descriptive and computational success in perturbative QED and the standard model, is certainly the enduring nature of its causality principle. In addition to experimental support through the validity of the Kramers–Kronig dispersion relations in high-energy collisions up to the shortest accessible distances, it is also the various unsuccessful theoretical attempts to construct viable nonlocal theories† which testify to the strength of this principle. Despite intense efforts and much talk, nobody has succeeded in constructing a viable *nonlocal* theory. The cutoff in Feynman-like integrals or in Euclidean functional integrals (which violate the prerequisites for continuability to real-time LQP) introduced by phenomenologists in order to combat the apparent 'bad' short-distance behaviour stemming from perturbative causality down to arbitrary small distances (which threaten the mathematical existence of models) are no substitute for a conceptual analysis of whether a viable nonlocal theory with an elementary length which maintains a particle interpretation is at all possible‡. Here, 'viable' is more than mere mathematical existence, it is meant in the physical sense of conceptual completeness. One requires that a theory should *contain its own physical interpretation* i.e. that one does not have to invent or borrow formulae from outside this theory as, for example, in phenomenological 'effective' QFT. In the latter case, most formulae linking the calculations with measurable quantities cannot be derived or justified but, as expected in a phenomenological approach, have to be taken from a more complete and fundamental framework. In addition, 'effective' indicates that objects with this prefix such as Lagrangians, actions etc should be dealt with using different rules such as those in renormalized perturbation theory. On the other hand, in a complete framework such as LQP, one cannot only derive (LSZ) scattering formulae which constitute an important aspect of particle interpretation, but one can also obtain the composition laws of charges, analytic and crossing properties of fields in particle states etc; in fact there is presently no important structural or epistomological property which the principles of LQP cannot address or account for. Only when it comes to the quantitative understanding of particle interaction processes does one have to resort to specific models, even though their full control is often very problematic as a result of the absence of systematic and reliable nonperturbative methods.

In contrast to statements one sometimes finds in the literature, there is no known nonlocal Poincaré covariant scheme, which guarantees the existence of a time-dependent (or its stationary reformulation) scattering formalism together with the analytic and crossing properties of matrix elements of the S-operator and formfactors of local fields, which therefore could be used in particle physics. Hence the importance of causality is also highlighted by the failure of nonlocal modifications and the conspicuous absence of physically viable alternatives. It is quite instructive to look briefly at some of the more prominent failed attempts.

In the 1950s there were already proposals to inject nonlocal aspects through extended interaction-vertices in Lorentz invariant Lagrangians. As mentioned before, this was motivated by the hope that a milder perturbative short-distance behaviour in correlation functions may be helpful for demonstrating the mathematical existence of the theory. It was soon realized, that if one pursues the effect of such modifications up to infinite order in perturbation theory, these nonlocal vertices would even wreck macrocausality so that the theory loses its physical interpretation altogether. A similar fate occurred to the later proposal of Lee and Wick [2] to

† The meaning of 'nonlocal' in this paper is not that of extended charged objects in a theory of local observables (example: semi-infinite string-like spatial extensions of anyons or plektons in $d = 1 + 2$ in order to support their Abelian/non-Abelian braid group statistics), but rather refers to hypothetical theories which have a fundamental cutoff or elementary length in their algebra of observables.

‡ A good antidote against speculations or light-hearted attitudes that, e.g., rotational invariant Euclidean cutoffs (or any other kind of cutoff which formally can be expected to maintain Lorentz covariance) could define a consistent nonlocal real-time theory, is to try to introduce one into one of the exactly solvable $d = 1 + 1$ factorizing models.

allow for complex (+ complex conjugate, in order to maintain hermiticity) poles in Feynman rules; it led to unacceptable time precursors [3]. In the last section we present some results on a new nonperturbative framework which incorporates and explains all the results obtained so far on explicit non-Lagrangian low-dimensional model constructions. The very concepts of this approach use causality and locality in a much more essential way than the various quantization approaches and, in addition, this method throws considerable doubt on the belief that the perturbative link between good short-distance behaviour and the existence of the theory has general validity.

Often the renormalization group ideas are used to justify a physical cutoff with the hope that by softening short-distance behaviour the model becomes mathematically better defined and managable. But physical principles should receive their limitation, as always happened in the past, from other more general principles and not from parameters into which one tries to 'dump' ones lack of knowledge about the mathematical existence of the theory within the presently known principles. A phenomenologically successful parameter with fixed computational prescriptions is, by itself, no substitute for a physical principle. Physical reality may unfold itself like an onion with infinitely many layers of ever more general physical principles tending towards the small, but it should still be possible to have a mathematically consistent theory in each layer which is faithful to the principles valid in that layer. This has been fully achieved for quantum mechanics (QM), but this goal was not reached in QFT due to a lack of nontrivial $d = 1 + 3$ models or structural arguments which could demonstrate that the requirements allow for nontrivial solutions. Even the recently emphasized duality between asymptotically small/large coupling parameters only resulted in the rephrasing of the problem to: *does there exist a QFT which possesses these two asymptotes.* The existence problem of interacting QFTs in $d = 1 + 3$, which persists to present times, sets QFT apart from any other physical theory such as QM, statistical mechanics or classical particle/field theories. In all of those cases one has explicit examples as well as proofs that the 'axioms' are consistent with nontrivial dynamics. In this context one should note that lattice theories define a different (mathematically easier) framework which, if suitably restricted, shares the fact with QFT that it is conceptually complete as far as the notion of particle excitations and their scattering theory (based on cluster properties as a substitute for the missing locality) is concerned. In fact, the correlation functions of lattice algebras are expected to converge towards those of QFT in an appropriately defined scaling limit. Despite some control of the extremely difficult scaling limits in certain special models such as the $d = 2$ Ising-like models, the relation between the two theories is still not largely understood.

Recently, there has been a more sophisticated attempt to go beyond the causal setting of LQP via the use of noncommutative space-time [4], based on spatial uncertainty relations following from a quasiclassical quantization interpretation of Einstein's field equation of general relativity and the assumed absence of very small black holes (similar uncertainty relations for the complete set of coordinates and momenta (i.e. for phase space) have been postulated on the basis of string theory [5]). These proposals, especially if they are backed up by uncertainty relations whose derivation is carried out in the spirit of Bohr–Rosenfeld as in [4], are not as easily dismissed as the two previous ones. Such attempts do not just try to graft cutoffs or elementary length onto the standard (Lagrangian, functional integral) local framework, but rather are receptive to more radical changes of the fundamentals of QFT. It is not easy to confront such speculative new ideas with LQP, because it is more difficult to physically interpret in such unusual frameworks than it is to rule out implanting cutoffs into the standard framework. Whereas it is easy to agree that sufficient intelligent noncommutative space-time proposals may serve as interesting tests for exploring the unknown territory beyond the reign of Einstein causality, they are still far from being models for the elusive 'quantum gravity', since they only replace the classical space-time indexing of nets with a noncommutative one. However, any step beyond the present causal framework must re-obtain Einstein causality as a limiting statement within some yet unknown new physical principle. Recently there have been many promises on the basis of string theory. But unfortunately string theory, even aside from the total lack of experimental motivation, has hardly added anything to conceptual problems despite its undeniable mathematical enrichments. In fact, in its present state it is mainly a loose set of calculational recipes which suffer from a very unfortunate preference of formalism over conceptual clarifications. Whereas LQP allows an intrinsic characterization (e.g. in terms of correlation functions or observable nets) independent of the way they have been manufactured (e.g. Lagrangian quantization, bootstrap-formfactor method in $d = 1 + 1$), string theory, in its more than 20-year existence, has not led to objects with an intrinsic meaning independent of the computational rules, in addition to its experimental invulnerability which it acquired after it changed interpretation from the old string theory of the dual model for strong interaction at laboratory energies to an alleged theory of quantum gravitation thus jumping 15 orders of magnitude. On the theoretical side, such fundamental questions as whether strings are localized objects in space-time (as the name seems to indicate) or if the name is a shorthand notation for specific spectral features have not been settled yet. Whereas, admittedly many of the popular formulations of QFT based on canonical or functional integral quantization also start from extrinsic formal requirements which in most cases cannot be maintained after renormalization†, there exist at least various intrinsic formulations.

Causality and locality are related in a profound way to the foundations of quantum theory in the spirit of von Neumann. In von Neumann's formulation, observables are represented by self-adjoint operators and measurements are compatible if the operators commute. The totality of all measurements which are relatively compatible with a given set (i.e. noncommutativity within each set is allowed) generate a subalgebra: the commutant $\mathcal{L}'$ of the given set of operators $\mathcal{L}$. In LQP, a conceptual framework which was not available to von Neumann, one is dealing with an isotonic 'net' of subalgebras (in most physically interesting cases von Neumann factors, i.e. with a trivial centre) $\mathcal{O} \to \mathcal{A}(\mathcal{O})$, such that, unlike QM, the spatial localization and the time duration of observables becomes an integral part of the formalism. *Causality gives an a priori information about the size of space-time $\mathcal{O}$-affiliated operator von Neumann algebras:*

$$\mathcal{A}(\mathcal{O})' \supset \mathcal{A}(\mathcal{O}'). \tag{1}$$

In words, the commutant of the totality of local observables localized in the space-time region, $\mathcal{O}$, contains the observables localized in its space-like complement (disjoint) $\mathcal{O}'$. In fact, in most cases the equality sign will hold in which case one calls this strengthened (maximal) form of causality 'Haag duality'‡:

$$\mathcal{A}(\mathcal{O})' = \mathcal{A}(\mathcal{O}'). \tag{2}$$

In other words, the space-like localized measurements are not only commensurable with the given observables in $\mathcal{O}$, but every measurement which is commensurable with all observables in $\mathcal{O}$ is necessarily localized in the causal complement $\mathcal{O}'$. Here we extended, for algebraic convenience, von Neumann's notion of observables to the whole complex von Neumann algebra generated by Hermitian operators localized in $\mathcal{O}$. If one starts the theory from a net indexed by compact regions $\mathcal{O}$ as double cones, then algebras associated with unbounded regions $\mathcal{O}'$ are

† Apart from some less interesting super-renormalizable models, the physically meaningful renormalizable objects (which are also the only ones with a chance of mathematical existence) are neither canonical nor representable by functional integrals, but still fulfil the property of Einstein causality together with certain spectral properties. The so-called 'causal perturbation theory' (see later) furnishes a more harmonious intrinsic formulation for which the initial requirements are also reflected in the results, and not only as a 'catalyzer' of the mind.

‡ See [1], and for more details on aspects related to my notes see [6].

defined as the von Neumann algebra generated by all $\mathcal{A}(\mathcal{O}_1)$ if $\mathcal{O}_1$ ranges over all net indices $\mathcal{O}_1 \subset \mathcal{O}'$.

Whereas the Einstein causality (1) allows a traditional formulation in terms of point-like fields $A(x)$ as

$$[A(x), A(y)] = 0 \qquad (x-y)^2 < 0 \tag{3}$$

Haag duality can only be formulated in the algebraic net setting of LQP. This aspect is shared by many important properties and results presented in this paper. LQP is much more than a teutonic pastime of reformulating properties of fields in terms of algebraic properties of nets, as one immediately realizes if one looks at Haag's book.

One can prove that Haag duality always holds after a suitable extension of the net to the so-called dual net $\mathcal{A}(\mathcal{O})^d$. The latter may be defined independent of locality, in terms of relative commutation properties, as

$$\mathcal{A}(\mathcal{O})^d := \bigcap_{\mathcal{O}_1, \mathcal{O}_1' \subset \mathcal{O}} \mathcal{A}(\mathcal{O}_1)'. \tag{4}$$

It is easy to check that the dual net is relatively local to the original net

$$A(\mathcal{O}_1) \subset (\mathcal{A}(\mathcal{O})^d)' \qquad \mathcal{O}_1 \subset \mathcal{O}'. \tag{5}$$

In fact, it is the maximal net relatively local to $\mathcal{A}(\mathcal{O})$. Repeating this process, one obtains $\mathcal{A}(\mathcal{O})^d \subset \mathcal{A}(\mathcal{O})^{dd}$ and $\mathcal{A}(\mathcal{O})^d = \mathcal{A}(\mathcal{O})^{ddd}$. Causality of the original net then means $\mathcal{A}(\mathcal{O}) \subset \mathcal{A}(\mathcal{O})^d$, and, therefore, also $\mathcal{A}(\mathcal{O})^{dd} \subset \mathcal{A}(\mathcal{O})^d = \mathcal{A}(\mathcal{O})^{ddd}$. It is customary to use the word locality instead of causality if one allows field algebras which involve fermionic structures. Local algebras retain all of the mathematical properties of observable algebras in that they contain no local annihilators. The extension by charged objects with braid group statistics (only possible in space-time dimension $d < 1 + 3$) may lead to algebras (acting in a larger Hilbert space) with weaker locality properties and the appearance of local annihilators. Such objects are called 'localizable' since they maintain their *relative locality* with respect to the neutral observable subalgebra. The causal disjoint of the region of relative commutation is the localization region of these charged operators. These considerations show that causality, locality and localization in LQP have a close relation to the notion of compatibility of measurements [7]. The fundamental reason for all such modifications in the interpretation of LQP versus QM is the different structure of local algebras: the vacuum is not a pure state with respect to any algebra which is contained in an $\mathcal{A}(\mathcal{O})$ with $\mathcal{O}'$ nonempty, and the sharply localized algebras $\mathcal{A}(\mathcal{O})$ do not admit any pure states at all! Since these fine points can only be appreciated with some more preparation, I will postpone their presentation. Note that the quantization approach to QFT based on the use of classical actions in Euclidean functional integrals (and the subsequent use of analytic continuation to get back to real space-time) is a global attempt to characterize vacuum expectation values of a would be theory. The classical locality in the sense of local polynomial expressions in fields and derivatives has no direct conceptual relation with the real time locality in the above sense; in fact the analytically continued 'fields' in the Euclidean points are extremely nonlocal relative to the real-time fields. Unlike in statistical mechanics, it does not make sense to restrict the Euclidean integration to localized configurations with local supports since this has nothing to do with the localization of real-time physics which is implemented via the restriction of states to localized subalgebras. Nevertheless, there are sufficient conditions under which the Euclidean correlation functions do permit one to define models of real-time QFT.

If the vacuum net is Haag dual, then all associated 'charged' nets share this property, unless the charges are non-Abelian; in which case the deviation from Haag duality is measured by the Jones index of the above inclusion, or in physical terms the statistical or quantum dimension. Even if the vacuum representation violates Haag duality, this indicates spontaneous symmetry breaking [7], i.e. not all internal symmetry algebraic automorphisms are spatially implementable. As already mentioned, in that case one can always maximize the local algebra to the dual algebras $\mathcal{A}^d(\mathcal{O})$ without destroying causality and without changing the Hilbert space and in this way Haag duality is restored ('essential duality'). This turns out to be related to the descent to the unbroken part of the symmetry which allows (since it is a subgroup) more invariants, i.e. more observables. Although these matters are good illustrations of the pivotal role of causality, we concentrate on the closely related modular properties of causal nets which make their appearance in the next section. QM does not know these concepts at all, trying to add them would mean leaving QM, since their realization requires infinite degrees of freedom and the virtual particle structure of the vacuum (together with the ensuing type III_1 von Neumann algebra structure of the local algebras which is remarkably different from that of quantum mechanical algebras).

Another structurally significant deviation is expected to result from the fact that the vacuum becomes a thermal state with respect to the local algebras $\mathcal{A}(\mathcal{O})$. There are two different mechanisms which generate thermal states: the coupling with a heat bath and the thermality through restriction or localization and the creation of horizons. The latter is in one class with the Hawking–Unruh mechanism; the difference being that in the localization situation the horizon is not classical, i.e. is not defined in terms of a differential geometric Killing generator of a symmetry transformation of the metric.

Since the algebras of the type $\mathcal{A}(\mathcal{O})$ do not possess pure states, the $\mathcal{O}/\mathcal{O}'$ situation is totally different from the tensor product factorization in terms of the quantization box inside/outside in QM. In order to return to a tensor product situation and be able to apply the concepts of entanglement and entropy, one has to do a sophisticated split which is only possible if one allows for a 'collar' (see later) between $\mathcal{O}$ and $\mathcal{O}'$. These considerations show that certain things which one takes for granted as properties of general QT actually lose their validity in LQP.

Since the thermal aspects of localization are analogous to those of black holes, there is no chance to directly measure such tiny effects. However, in conceptual problems, e.g. the question of if and how not only classical relativistic field theory but also QFT excludes superluminal velocities, these subtle differences play a crucial role. Imposing the usual algebraic structure of QM onto the theory of photons will lead to nonsensical results. Most sensational theoretical observations on causality violations which are not already incorrect on a classical level suffer from incorrect tacit assumptions. We urge the reader to read [18] and also look at the source for that rebuttal.

Historically, the first conceptually clear definition of localization of a relativistic wavefunction was given by Newton and Wigner [8] who adapted Born's x-space probability interpretation to the Wigner relativistic particle theory. Apparently, the result that there is no exact satisfactory relativistic localization (but only one sufficient for all practical purposes), disappointed Wigner so much that he became distrustful of the consistency of QFT in particle physics altogether (private communication by R Haag). Whereas we know that this distrust was unjustified, at the same time we should acknowledge Wigner's stubborn insistence on the importance of the locality concept as an indispensable particle physics requirement in addition to the positive-energy property and irreducibility of his representations theory. Modular localization of subspaces of the Hilbert space and of subalgebras, on the other hand, are not related to the Born probability interpretation. Rather, modular localized state vectors pre-empt the existence of causally localized observables and have no counterpart at all in N-particle QM. As is explained later, modular localization may serve as a starting point for the construction of interacting nonperturbative LQPs [6, 10]†. It is worthwhile emphasizing that sharper localization of local algebras in LQP is not defined in terms of smaller support properties of classical smearing functions of smeared fields but rather in terms of the intersection of algebras; although in many cases such as CCR- or CAR-algebras (or more generally Wightman fields) the algebraic formulation (1) can be reduced to this more classical concept.

Since the modular structure is related in a fundamental way to thermal behaviour, it is not surprising that the issue of thermality is also related with localization. In fact, as mentioned before, there are two manifestations of thermality, the standard heat-bath thermal behaviour which is described by the Gibbs formula (or after having performed the thermodynamic limit by the KMS condition), and thermality caused by localization with either classical bifurcated Killing-horizons as in black holes [9], or in a purely quantum manner as the boundary of the Minkowski space wedges or double cones. In the latter case the KMS state has no natural limiting description in terms of a Gibbs trace-formula (which only applies to type I and II, but not to type-III von Neumann algebras), a fact which is also related to the fact that the Hamiltonian (of the ground state problem) is bounded from below, whereas, e.g., the Lorentz boost (the modular operator of the wedge algebra in the vacuum state), is not [10]. In [11] the reader also finds a discussion of localization and cluster properties in a heat-bath thermal state. In this paper we do not enter these interesting thermal aspects. Recent results indicate that the division between heat bath- and localization-thermality may not be as sharp as it appears at first sight [55].

2. Locality and free particles

The best way to make the pivotal nature of causality manifest is to enter QFT via Wigner's group theoretical characterization of particles by irreducible positive-energy representations with good localization properties. It is well known that the Wigner wavefunctions ψ of massive spin-s particles have $2s + 1$ components and (differently from covariant fields) transform in a manifestly unitary but p-dependent way:

$$(U(\Lambda)\psi_W)(p) = R(\Lambda, p) \cdot \psi_W(\Lambda^{-1}p). \tag{6}$$

The transition to covariant wavefunction and fields is performed with the help of intertwiners $u(p, s_3)$, resp. the rectangular matrix $U(p)$, constructed from their $2s + 1$ column vectors of length $(2A + 1) \cdot (2B + 1)$:

$$U(p)D^{(s)}(R()) = D^{(A,B)}(\Lambda)U(\Lambda^{-1}p) \tag{7}$$

i.e., within Wigner's Poincaré group positive-energy representation theory one can intertwine the rotations (with the p-dependent Wigner R-matrix) with the (dotted and undotted) finite-dimensional spinor representations $D^{(A,B)}$. Since the $D^{(s)}$ representation of the rotations is 'pseudo-real', there exists another intertwiner matrix $V(p)$ which is 'charge-conjugate' to $U(p)$. To each of the infinitely many intertwiner systems (the only restriction on A, B for a given physical spin s is $|A - B| \leqslant s \leqslant |A + B|$) one has a local field obeying the spin-statistics connection:

$$\psi^{(A,B)}(x) = \frac{1}{(2\pi)^{\frac{3}{2}}} \int \left(\mathrm{e}^{-\mathrm{i}px} \sum u(p, s_3)a(p, s_3) + \mathrm{e}^{\mathrm{i}px} \sum v(p, s_3)b^*(p, s_3) \right) \frac{\mathrm{d}^3 p}{2p_0} \tag{8}$$

† In fact, the good modular localization properties of positive-energy properties, with the exception of Wigner's infinite component 'continuous spin' representations, are guaranteed. Only in the infinite component case it is not possible to come from the wedge localization to the space-like cone localization, which is the coarsest localization from which one can still obtain a Wigner particle interpretation.

where a, b are the (creation) annihilation operators associated with the Fock space enlargement of the Wigner representation space and hence independent of the choice of intertwiners. All the different fields are describing the same (m, s) particle physics and live in the same Fock space. They constitute only the linear part of a huge (Borchers) equivalence class of fields. For free fields, this equivalence class contains, in addition, all Wick-monomials, and it is well known that they are indispensible for introducing perturbative interactions. The above different ψ' can be mutually solved:

$$\psi^{(A',B')}(x) = M_{(A,B)}^{(A',B')}(\partial)\psi^{(A,B)}(x) \tag{9}$$

where $M_{(A,B)}^{(A',B')}(\partial)$ is a rectangular matrix (matrix indices supressed) involving ∂_μ derivatives.

Explicit formulae can be found in the first volume of [12]. Among the infinitely many possibilities, essentially only one is 'Lagrangian', i.e. can be used in a quantization approach starting from a classical Hamiltonian principle. The other descriptions are equally physically acceptable, since there is no quantization principle which forces one to perform quantum physics through a classical parallelism with the Lagrangian formalism. In fact, they describe the same physics in the form of a different 'field coordinatization'.

Indeed for LQP, point-like fields (8) are like coordinates in differential geometry; the different covariant realizations of the Wigner representations correspond to linear coordinate changes, whereas the transition from one cyclically acting field in the Borchers equivalence class to another constitutes the nonlinear part. Although it may be sometimes convenient to use them but structural theorems on charge-carrying fields (classification of statistics, including braid-group statistics for low-dimensional charge carriers, TCP ...) and internal symmetries (symmetries and their spontaneous breaking a la Nambu–Goldstone, the Schwinger–Higgs screening mechanism ...) are best done in terms of the properties of the net:

$$\mathcal{O} \to \mathcal{A}(\mathcal{O}). \tag{10}$$

The causality and spectral properties of these nets constitute the physical backbone of LQP. The notion 'local' is then extended to all boson and fermion fields, because they allow an unrestricted iterative application to the vacuum without encountering local annihilators, and therefore such an extension preserves the important properties of the original observables. More general charge-carrying fields which extend the above local (bosonic or fermionic) net are called 'localizable' (with respect to the observables). In particular, plektonic (braid-group statistics) $d = 1{+}2$ dimensional fields can never have a Fock space structure and always locally annihilate charge sectors when the operator domain does not match the range of the charge sector of the state vector. Although such fields (as some fields used in gauge theory) necessarily have a semi-infinite (space-like) string-like extension, these charge carriers are associated with a local net of observables i.e. they do not bring in an aspect of elementary length or any other restriction of the causality principle. A genuinely nonlocal theory would *violate causality in its observable algebra*; as long as the theory admits a causal observable algebra there is no elementary length, independently of the possibly extended nature of charged operators. In other words, extended operators which transfer charges and communicate between different representations of the observables are permitted as long as their commutation relations relative to the observables reflect their spatial extension in the previously mentioned sense.

It is important to note that the Wigner free fields have operator dimensions (referring to the short-distance power behaviour) which increase with spin: $\dim \psi_{(s=0)} = 1$, $\dim \psi_{(s=\frac{1}{2})} = \frac{3}{2}$, $\dim \psi_{(s=1)} \geqslant 2$. This is the deeper reason why the incorporation of interacting theories into the scheme of causal renormalized perturbation requires special (BRS) cohomological tricks for $s \geqslant 1$ (the LQP version of gauge theories, see the next section).

The Wigner approach for $(m = 0, h \geqslant 1)$ leads to a more restricted class of intertwiners, since many representations (e.g. the $D^{(\frac{1}{2},\frac{1}{2})}$ vector representation), as a result of the different nature of the 'little' group, *cannot* be intertwined with the physical photon ($m = 0, h = 1$) of the Wigner representation theory. In fact, the range of dotted/undotted indices in (8) is restricted according to $h = \pm|A - B|$ [12]. There are two methods to overcome this restriction; one physical way of introducing a semi-infinite space-like localized vector potential $A_\mu(x, n)$ depending on a space-like string direction n into the Wigner photon space, or the extension by ghost fields (indefinite metric or different star-operation) formalism which keeps the formal Lorentz-covariance (together with the point-like nature) in the form of 'pseudo-unitarity' representations. Whereas the first method is physically deeper and more promising, the second one is the only one which is compatible with the presently known formalism of renormalized perturbation theory. The latter does not care whether the locality is formal instead of physical or whether the boost transformations are pseudo-unitary instead of unitary, but the interpretation does.

The remaining positive-energy representations are Wigner's famous 'continuous spin' representations which are infinite components (infinite-dimensional representations of the massless 'little group'). They are usually dismissed by saying that nature does not make use of them. Apart from the fact that a theoretician should not argue in this way (and in fact does not, if it comes to supersymmetry), the dismissal is probably founded on the naive identification of irreducible positive-energy representation with physical particles. This ignores the fact that particles should be described by states, which in addition to forming irreducible positive-energy representations, must also have good localization properties. The modular localization method below reveals that any positive-energy representation can be localized in wedges. For all positive-energy representations with finite spin/helicity the localization can be sharpened; for the $m = 0$ continuous spin representations, however, the same methods are inconclusive. It is doubtful that they admit a sharper localization, needed for particle interpretation including scattering, and this may cause their disqualification as candidates for physical particles on the theoretical side. There are also many useful particle-like objects or states which are not described by ($m, s =$ semi-integer) Wigner representations as, e.g., infraparticles (electron with photon cloud), ultraparticles, quarks, etc [13]. The borderline between physical particle and other weakly localizable objects is the string-like (or, more appropriately, space-like-cone) localization. This localization is still sufficient to derive scattering theory, and on the other hand, it follows from the existence of field theoretic charge sectors which fulfil the mass gap assumption [1]. Operators with braid-group commutation relations in $d = 1 + 2$ which have one-particle components with mass gaps, are necessarily string-like and lead to anyons (Abelian, spin arbitrary) or plektons (non-Abelian, spin quantized). Therefore, compactly (e.g. double-cone) localizable fields and particles in $d = 1 + 2$ are only consistent with the permutation group statistics which is a special case of braid-group statistics.

If fields are analogous to coordinates in differential geometry, there should be a way to at least construct interaction-free nets directly, without ever using free fields. The idea behind this is to characterize wedge localized real subspaces in Wigner space with the help of modular operators (instead of Cauchy initial value data). For simplicity assume integer spin self-conjugate bosons and define a real subspace $H_R(W_{st})$ of H_{Wigner} as:

$$\begin{aligned} &H_R(W_{st}) = \text{closure of real lin. comb.}\{\psi|s\psi = \psi\} \\ &s \equiv j\delta^{\frac{1}{2}} \qquad s^2 = 1. \end{aligned} \tag{11}$$

The notation is as follows: $\delta^{i\tau} := U(\Lambda_{x,t}(2\pi\tau))$ is the Lorentz boost in the x–t direction associated with the standard x–t wedge $W_{st} := \{x \in \mathbb{R}^4; x_1 > |x_0|\}$, and $j = \theta \cdot \mathrm{rot}_x(\chi = \pi)$ is, apart from a π-rotation around the x-axis, the antiunitary TCP transformation θ acting on the Wigner one-particle space, which for nonself-conjugate particles consists of a direct sum of the particle and antiparticle space. The unbounded $\delta^{\frac{1}{2}} > 0$ is defined by functional calculus from δ^{it}

and has a domain consisting of boundary values of an analytically continuable $2s+1$ component wavefunction which have the momentum space rapidity ($p_0 = m\cosh\theta$, $p_x = m\sinh\theta$) analyticity in the strip $-\pi < \operatorname{Im}\theta < 0$. The unbounded s inherits the densely defined domain from $\delta^{\frac{1}{2}}$ and the antilinearity from j. The best way to describe this real Hilbert space of wedge-localized functions is to say that they are strip-analytic and fulfil the Schwartz reflection principle around the line $\operatorname{Im} z = -\frac{i\pi}{2}$. In the case of antiparticles $\neq$ particles one must double the number of components and use the full charge conjugated wavefunctions in the reflection principle instead of just the complex conjugate. This is closely related to the crossing 'symmetry' (it is not a symmetry in the standard operational sense of QT) in interacting systems, about which we will have more to say later. The involutive property $s^2 = 1$ on this domain, in mathematical notation $s^2 \subset 1$, is a consequence of this definition. Such unbounded (but yet involutive) operators do not occur in any other area of mathematical physics and are therefore not treated in books on mathematical methods. In fact they seem to be characteristic of the Tomita–Takesaki modular theory. It is precisely the combination of unboundedness and involutiveness which is responsible for the emergence of localization and geometrical properties from domain properties of quantum physical operators. The real closed subspace may be used to define a dense wedge-localization space† $H(W_{st}) \equiv H_R(W_{st}) + iH_R(W_{st})$ on which the operator s acts as:

$$s(h + ih) = h - ih. \tag{12}$$

$H_R(W_{st})$ is 'standard' i.e.

$$\begin{aligned} &H_R(W_{st}) \cap iH_R(W_{st}) = \{0\} \\ &H(W_{st}) \equiv H_R(W_{st}) + iH_R(W_{st}) \qquad \text{is dense in} \quad H_{Wigner}. \end{aligned} \tag{13}$$

The natural localization topology is the graph norm of s. It is somewhat unusual and treacherous that the formula for s looks so universal and the differences in the localization for different wedges $W_\Lambda := \Lambda W_{st}$, $W_a := T(a)W_{st}$ is solely encoded in the domain of definition of $s_{(\Lambda,a)}$ (i.e. only where and not how it acts) which he usually considers to be a fine and somewhat irrelevant technical point. For positive-energy representations the geometric inclusion $W_a := T(a)W_{st} \subset W_{st}$, $a \in W_{st}$ (translating wedges into themselves) implies the proper inclusion (D Guido 1996 private communication) $H_R(W_a) \subset H_R(W_{st})$, in fact the geometric inclusion properties are equivalent to the positive spectrum condition. For the understanding of the latter claim one has to decompose the space-like a into two light-like components $a_\pm$ for which one takes, of course, the two light-like vectors by which the wedge W_{st} is generated. Different from space-like translations, these light-like translations have a positive generator.

Having constructed a net of wedge-localized real subspaces $H_R(W)$, one may move ahead and introduce compactly localized spaces $H_R(\mathcal{O})$ through intersections of wedges

$$H_R(\mathcal{O}) = \cap_{W\supset\mathcal{O}} H_R(W). \tag{14}$$

In order to insure the nontriviality of these intersections, one needs to restrict the positive-energy representations to those with a finite-dimensional representation of the Wigner 'little group' which amounts to (half)integer spin/helicity. In this way one obtains, e.g., the net of double cones; a direct construction of the associated modular objects is more difficult because the modular group behaves 'geometrically' (i.e. as a diffeomorphism of Minkowski space) only asymptotically close to the 'horizon' (the boundary of the causal closure) of the region. Note that in order to define these localization spaces, we did not use any u, v intertwiners. If

† A change of sign in the definition of $H_R(W)$ would not change the dense complex localization space (which is a Hibert space in the graph s-norm).

we had, the present intrinsic concept of localization would have been lost and we would have been back at x-space properties of covariant wavefunctions or pointlike fields, i.e. those field coordinatizations which destroyed the uniticity. The size of localization is contained in certain Payley–Wiener–Schwartz type of bounds in imaginary momentum or rapidity directions.

The last step to the nets consists (say for the case of integer spin) in the application of the Weyl functor which maps real subspaces into the von Neumann subalgebras of a net:

$$\begin{aligned} & H_R(\mathcal{O}) \stackrel{\mathcal{F}}{\rightarrow} \mathcal{A}(\mathcal{O}) \\ & \mathcal{A}(\mathcal{O}) = \mathrm{alg}\{W(f)|f \in H_R(\mathcal{O})\} \\ & W(f) = \mathrm{e}^{\mathrm{i}(a^*(f_1)+\mathrm{h.c.})+\mathrm{i}(b^*(f_2)+\mathrm{h.c.})} \qquad f = (f_1, f_2) \end{aligned} \tag{15}$$

where $b^\#$, $a^\#$ stand for (anti)particle Wigner creation and annihilation operators. The functor $\mathcal{F}$ is 'orthocomplemented' which means that the symplectic or (by multiplication with i) real orthogonal complement of a real subspace is mapped into the von Neumann algebraic commutant. The images J, Δ^{it}, S of j, δ^{it}, s under $\mathcal{F}$ are the modular objects of the algebraic version of the Tomita Takesaki modular theory for the special case of the pair $(A(W_{st}), \Omega)$ of wedge algebra and vacuum vector†.

The general theory says that for a von Neumann algebra $\mathcal{A}$ with a cyclic and separating vector Ω, the definition:

$$SA\Omega = A^*\Omega \qquad A \in \mathcal{A} \tag{16}$$

introduces a closable operator, whose polar decomposition;

$$S = J\Delta^{\frac{1}{2}} \tag{17}$$

defines a unitary Δ^{it} and an antiunitary involution J which are of fundamental significance for the pair $(\mathcal{A}, \Omega)$. The operator Δ^{it} defines the 'modular' automorphism σ_t of $\mathcal{A}$ (a kind of generalized Hamiltonian) with respect to Ω and J the modular involution j (a kind of generalized TCP reflection):

$$\begin{aligned} & \sigma_t(\mathcal{A}) = \mathcal{A} \qquad \sigma_t(A) \equiv \Delta^{it} A \Delta^{-it} \\ & j(\mathcal{A}) = \mathcal{A}' \qquad j(A) \equiv JAJ. \end{aligned} \tag{18}$$

This basic theorem was stated and proved by Tomita with significant improvements due to Takesaki [14]. In the context of thermal quantum physics it received an important independent contribution in the form of the KMS condition from Haag Hugenholz and Winnink (HHW); whereas Kubo, Martin and Schwinger only used this analytic condition in order to avoid the calculation of traces, HHW elevates this property to one of the most important conceptual tools related to the stability of states and to the second thermodynamical law [1]. Its relevance for localization in QFT was first seen in full generality by Bisognano and Wichmann [17] and the thermal aspects of (wedge) localization (the Hawking–Unruh connection) were first stressed by Sewell [9]. Although we explained the construction of free nets only for bosons, the formalism adapts easily to fermions. Fermions are pre-empted in the modular localization of the Wigner theory by the appearance of a mismatch between the geometrical opposite of $H_R(W)$ obtained by a 180° rotation, and its symplectic or real orthogonal complement. This leads to a modification of the Tomita involution in the form of an additional twist which can be shown to pre-empt the Fermi-statistics already on the one-particle level. Our inverse use of the Bisognano–Wichmann idea for the purpose of direct net construction which we exemplified for free theories in arbitrary space-time dimensions can be generalized to interacting theories

† A construction of the free net without using modular localization methods can be found in [16]. It is, however, the modular method which extends to the interacting case.

with the mathematical control being restricted presently to $d = 1 + 1$. Some of these results will be presented in the last section.

Already in the very early development of algebraic QFT [15] the nature of the single local von Neumann algebras became an interesting issue. Although it was fairly easy (and expected) to see that i.e. wedge- or double-cone-localized algebras are von Neumann factors (in analogy to the tensor product factorization of standard QT under formation of subsystems, it took the ingenuity of Araki to realize that these factors were of type III (more precisely hyperfinite type III_1 as we know nowadays, thanks to the profound contributions of Connes and Haagerup), at that time still an exotic mathematical structure. Hyperfiniteness was expected from a physical point of view, since approximatability as limits of finite systems (matrix algebras) harmonizes very well with the idea of thermodynamic + scaling limits of lattice approximations. A surprise was the type III_1 nature which, as already mentioned, implies the absence of pure states (in fact all projectors are Murray von Neumann equivalent to the identity operator) on such algebras; this property in some way anticipated the thermal aspect (Hawking–Unruh) of localization. Overlooking this fact which makes local algebras significantly different from algebraic aspects of QM, it is easy to make conceptual mistakes which could, e.g., suggest an apparent breakdown of causal propagation. For the discussion of such a kind of error and its correction see [18], as already mentioned in the introduction. If one simply grafts concepts of QM onto the causality structure of LQP (e.g. quantum mechanical tunnelling, structure of states) without deriving them in LQP, one runs the risk of wrong conclusions about, e.g., the possibility of superluminal velocities.

Let me, at the end of this section mention two more structural properties, intimately linked to causality, which distinguish LQP rather sharply from QM. One is the Reeh–Schlieder property:

$$\begin{aligned} &\overline{\mathcal{P}(\mathcal{O})\Omega} = H \qquad \text{cyclicity of } \Omega \\ &A \in \mathcal{P}(\mathcal{O}) \qquad A\Omega = 0 \Longrightarrow A = 0 \quad \text{i.e. } \Omega \text{ separating} \end{aligned} \tag{19}$$

which either holds for the polynomial algebras of fields or for operator algebras $\mathcal{A}(\mathcal{O})$. The first property, namely the denseness of states created from the vacuum by operators from arbitrarily small localization regions (a state describing a particle behind the moon† and an antiparticle on the earth can be approximated inside a laboratory of arbitrary small size and duration) is totally unexpected from the global viewpoint of general QT. In the algebraic $\mathcal{A}(\mathcal{O})$ formulation this can be shown to be dual to the second one (in the sense of passing to the commutant), in which case the cyclicity passes to the separating property of Ω with respect to $\mathcal{A}(\mathcal{O}')$.

Of course the claim that somebody causally separated from us may provide us with a dense set of states is somewhat unusual if one thinks of the factorization properties of ordinary Schrödinger-QT. The large enough commutant required by the latter property is guaranteed by causality (the existence of a nontrivial $\mathcal{O}'$) and shows that causality is again responsible for the unexpected property. If the naive interpretation of cyclicity/separability in the Reeh–Schlieder theorem leaves us with a feeling of science fiction (and also has attracted a lot of attention in philosophical quarters), the challenge for a theoretical physicist is to find an argument why, for all practical purposes, the situation nevertheless remains similar to QM. This amounts to the fruitful question: which among the dense set of localized states can be really produced with a controllable expenditure (of energy)? In QM the asking of this question is not necessary, since the localization at a given time via support properties of wavefunctions leads to a tensor product factorization of inside/outside so that the inside state vectors are automatically never dense in

† This weird aspect should not be held against QFT but rather be taken as indicating that localization by a piece of hardware in a laboratory is also limited by an arbitrary large but finite energy, i.e. is a 'phase space localization' (see the subsequent discussion). In QM one obtains genuine localized subspaces without energy limitations.

the whole space. Later we will see that most of the very important physical and geometrical informations are encoded into features of dense domains, in fact the aforementioned modular theory explains such relations. For the case at hand the reconciliation of the paradoxical aspect of the Reeh–Schlieder theorem with common sense has led to the discovery of the physical relevance of *localization with respect to phase space in LQP*, i.e. the understanding of the *size of degrees of freedom* in the set:

$$\begin{aligned} &P_E\mathcal{A}(\mathcal{O})\Omega \text{ is compact} \\ &\mathrm{e}^{-\beta \boldsymbol{H}}\mathcal{A}(\mathcal{O})\Omega \text{ is nuclear} \qquad \boldsymbol{H} = \int E\,\mathrm{d}P_E. \end{aligned} \tag{20}$$

The first property was introduced long ago by Haag and Swieca [1] whereas the second statement (and similar nuclearity statements involving modular operators of local regions instead of the global Hamiltonian) which is more informative and easier to use, is a later result of Buchholz and Wichmann [19]. It should be emphasized that the LQP degrees of freedom counting of Haag–Swieca, which gives an infinte but still compact set of localized states is different from the finiteness of degrees of freedom per phase space volume in QM, a fact often overlooked in the present day's string theoretic degree of freedom counting. The difference to the case of QM is diminished if one uses, instead of a strict energy cutoff, a Gibbs damping factor $\mathrm{e}^{-\beta H}$ as above. In this case the map $\mathcal{A}(\mathcal{O}) \to \mathrm{e}^{-\beta H}\mathcal{A}(\mathcal{O})\Omega$ is 'nuclear' if the degrees of freedom are not too accumulative (which then would cause the existence of a maximal Hagedorn temperature). The nuclearity assures that a QFT, which was given in terms of its vacuum representation, also exists in a thermal state. An associated nuclearity index turns out to be the counterpart of the quantum mechanical Gibbs partition function [1] and behaves in an entirely analogous way.

The peculiarities of the above Haag–Swieca degrees of freedom counting are very much related to one of the oldest 'exotic' and at the same time characteristic aspects of QFT, namely vacuum polarization. As discovered by Heisenberg, the partial charge:

$$Q_V = \int_V j_0(x)\,\mathrm{d}^3x = \infty \tag{21}$$

diverges as a result of uncontrolled vacuum fluctuations near the boundary. For the free-field current it is easy to see that a better definition involving test functions, which takes into account the fact that the current is a four-dimensional operator-valued distribution and has no restriction to equal times, leads to a finite expression. The algebraic counterpart is the so-called 'split property', namely the statement [1] that if one leaves between say the double-cone (the inside of a 'relativistic box') observable algebra $\mathcal{A}(\mathcal{O})$ and its causal disjoint (its relativistic outside) $\mathcal{A}(\mathcal{O}')$ a 'collar' $\mathcal{O}_1 \cap \mathcal{O}$, i.e.

$$\mathcal{A}(\mathcal{O}) \subset \mathcal{A}(\mathcal{O}_1) \qquad \mathcal{O} \ll \mathcal{O}_1 \qquad \text{properly} \tag{22}$$

then it is possible to construct in a canonical way a type I tensor factor $\mathcal{N}$ which extends in a 'fuzzy' manner into the collar $\mathcal{A}(\mathcal{O})' \cap \mathcal{A}(\mathcal{O}_1)$, i.e. $\mathcal{A}(\mathcal{O}) \subset \mathcal{N} \subset \mathcal{A}(\mathcal{O}_1)$. With respect to $\mathcal{N}$ the vacuum state factorizes, i.e. as in QM there are no vacuum fluctuations for the 'smoothened' operators in $\mathcal{N}$. The algebraic analogue of Heisenberg's smoothening of the boundary is the construction of a factorization of the vacuum with respect to a suitably constructed type-I factor algebra which uses the collar extension of $\mathcal{A}(\mathcal{O})$. It turns out that there is a canonical, mathematically distinguished factorization, which lends itself to define a natural 'localizing map' Φ which gives valuable insight into an intrinsic LQP version of Noether's theorem [1], i.e. one which does not rely on any parallelism to classical structures, as is the case with quantization. It is this 'split inclusion' which allows us to bring back the familiar structure of QM since type-I factors allow for pure states, tensor product factorization, entanglement and all the other properties at the heart of quantum theory and the measurement process. However, despite this partial return to structures known from QT, the vacuum retains its thermal properties even with respect to $\mathcal{N}$.

There are also interesting 'folklore theorems' i.e. statements which are mostly taken for granted, but for which yet no rigorous argument exists (but also no counter-example). One is the statement of 'nuclear democracy'. In the context of LQP it states that an operator from a (without loss of generality) double-cone algebra $A \in \mathcal{A}(\mathcal{O})$ or a point-like field couples to all states to which the superselection rules allow a nonvanishing matrix element. In particular we expect:

$$\langle \varphi^{in} | A | \psi^{in} \rangle \neq 0 \tag{23}$$

if the (say incoming) multiparticle state vector φ^{in} lies in the same charge superselection sector as $A|\psi^{in}\rangle$, i.e. 'everything communicates with everything' as long as the charges match†. A special case is the phenomenon of vacuum or better one-particle polarization through interaction i.e. the idea that there may be no interacting local operator $A \in \mathcal{A}(\mathcal{O})$ at all such that $A\Omega$ is in the one-particle space without additional $p\bar{p}$ contributions. In order to suppress this $p\bar{p}$ polarization cloud in state vectors of interacting theories, one has to allow at least a semi-infinite localization region as the wedge region. For any compact region, or even for those noncompact regions which are a tiny bit smaller than wedges, the infinite particle clouds and the field point of view take over. The polarization cloud content of a state vector $A\Omega$ with $A \in \mathcal{A}(\mathcal{O})$ is intimately related to the modular objects of $(\mathcal{A}(\mathcal{O}), \Omega)$. If one could back up these expectations (based on model observations) by rigorous theorems, one would have achieved an intrinsic understanding of interactions. Section 5 gives a brief account on what is presently known about the modular construction of interacting nets.

3. Renormalized perturbation, problems with $s \geqslant 1$

Following Tomonaga, Feynman, Schwinger and the other pioneers of perturbative renormalization, interactions are traditionally introduced through one of the various forms of quantization (canonical, path integral, etc).

The method which brings out the pivotal role of causality in the most explicit way is, however, the so-called 'causal perturbation method' which goes back to Stückelberg and Bogoliubov [20] which was formulated as a finite iteration method within the principles of LQP without reference to quantization by Epstein and Glaser [21]. Some refinements of that method, notably related to curved space-time and gauge theories, have been recently added to by [22, 23]. Also Weinberg's more formal derivation of Feynman rules for arbitrary spin [12] is somewhat in the spirit of causal perturbations.

It is a conceptual weakness of any quantization approach that contrary to QM, where this can be given a rigorous meaning, quantization in field theory remains more on the intuitive artistic side. Only for a so-called super-renormalizable interaction is the assumed canonical or functional Feynman–Kac quantization structure also reflected in the renormalized result; in all other cases it only serves as a vehicle which activates physicists thought and does not survive the renormalization procedure: i.e. with the mentioned exception no renormalized result fulfils canonical commutation relations or functional integral representations, rather the only surviving structure is causality/locality. This artistic rather than mathematical aspect pervades the standard textbook formulation of QFT. Such a state of affairs is acceptable, as long as one remains aware that (what I summarily call) the Lagrangian quantization is basically an

† This forces the substitution of the QM hierarchical concept of bound state particles in favour of charge fusion in LQP, which in turn means 'nuclear democracy' between particles.

efficient chain of formal manipulations and tricks which lead from slightly wrong assumptions after some repair to the correct perturbative results. Whereas the canonical structure and the functional integral representation cannot be upheld, the physical causality properties do survive the necessary repair procedure, better known under the name of renormalization.

In order to rescue the canonical or functional structures at any cost, physicists sometimes resort to imagining the existence of physical cutoffs or regulators and use the euphemism 'cutoff canonical variables or cutoff functional representations' without confronting those conceptual problems of noncausal/nonlocal theories mentioned in the introduction. In this way of thinking, the infinities of the unrenormalized theory relative to the renormalized, are sometimes attributed more physical significance than just indicating the necessity of repairing a slightly incorrect classical starting point (the classical Poincaré–Lorentz particle models within a classical field theory, instead of the Wigner-particle picture), which would be avoided in the causal perturbative approach.

To be fair, these conceptual drawbacks of the quantization artistry are partially offset by the efficiency of renormalizing away infinities through Feynman rules. Even if, e.g., Schwinger's finite-split-point method for the nonlinear terms in field equations may be conceptually cleaner, because one never meets a manifest infinity (as long as one does not interchange short distance limits with the other operations), the method is harder to systematize and practically less efficient compared with Feynman's method of confronting infinities or ad hoc cutoffs.

Different from the quantization + repair of infinities, LQP only uses those physical assumptions which are also genuinely reflected in the results (causality, spectral properties, modular structure of local algebras etc). The principles are the same principles as standard QFT but it does so in a more conscientious way. In such an approach the short-distance properties of individual fields are, apart from perturbation theory (infinitesimal deformations around free fields), less tightly connected with the existence of the model. We will come back to this important point in the nonperturbative section 5. In the following we will illustrate the strength of the LQP point of view in perturbation theory. The renormalized results are, of course, the same as in the functional approach, but the derivation and the guiding physical ideas differ in an interesting way.

In causal perturbation theory, which may be considered as a particular form of perturbative LQP, the interaction is implemented by locally coupling the free fields (any choice possible, ψ does not have to be Lagrangian!) by an L-invariant sum over Wick monomials $W_i(x)$ and one defines the following formal transition operator in Fock space†:

$$\begin{aligned} &S(g,h) = T\mathrm{e}^{\mathrm{i}\int\{g(x)W(x)+h(x)\psi(x)\}\,\mathrm{d}^4x} \\ &\tilde{C} \subset \operatorname{supp} g \subset C \\ &g_i = \text{const in } \tilde{C} \end{aligned} \tag{24}$$

where $W(x) = \sum g_i W_i(x)$ and $C, \tilde{C}$ are large double-cone regions. In the following we specialize to one field and one coupling for simplicity of notation (the notation for the general case with several fields and monomials we leave to the reader). Already without the time-ordering T, the operator exponential is a mathematically delicate object since the smeared Wick-powers beyond the second are not essentially selfadjoint on their natural domains. With the time ordering it is more serious: apart from certain W with low operator dimensions (a situation which cannot occur in $d = 1 + 3$ dimensions), there is no operator functional $S(g)$ in Fock space for which a mathematical control has been achieved (no solution of the 'Bogoliubov

† There is no compelling physical reason besides the historical success in QED and the analogy with QM why outside of deformation of free fields the introduction of interactions should follow this pattern. The existence of perturbation theory in the sense of a deformation theory has in general no bearing on the existence of an associated nonpertubative version.

axiomatics' in $d = 1 + 3$). Causal perturbation theory does not attempt to make sense of $S(g)$ but only of its nth order power series term in g. Therefore one proceeds along the following two lines:

- *Extraction of general causality properties for $S(g)$ and related operators (the 'Bogoliubov axiomatics').* The basic causality in the time-ordered formalism is:

$$\begin{aligned} &T(\psi(x_1)\ldots\psi(x_n)) = T(\psi(x_1))\ldots\psi(x_k))\cdot T(\psi(x_{k+1})\ldots\psi(x_n)) \\ &\text{if} \quad x_j \notin x_i + \bar{V}_+ \quad i = 1,\ldots,k \quad j = k+1,\ldots,n. \end{aligned} \tag{25}$$

  For the purpose of (formally) extracting a causal net it is helpful to reformulate this property in terms of another relative transition operator:

$$\begin{aligned} &V(g,h) \equiv S(g,h=0)^{-1}S(g,h) \\ &\text{causality}: V(g,h_1+h_2) = V(g,h_1)V(g,h_2) \\ &\text{if} \quad \operatorname{supp} h_1 \notin \operatorname{supp} h_2 + \bar{V}_+. \end{aligned} \tag{26}$$

  With the local algebras being now defined as (the notation alg includes the von Neumann closure):

$$\mathcal{A}_g(\mathcal{O}) \equiv \operatorname{alg}\{V(g,h), \operatorname{supp} h \subset \mathcal{O}\}. \tag{27}$$

  In fact, a change of the coupling strength g outside C (see (24)) does not change the net $\mathcal{A}_g(\mathcal{O})$ for $\mathcal{O}$ inside $\tilde{C}$, except for a common unitary (the nets are isomorphic, i.e. considered to be identical):

$$\begin{aligned} &V(g+\delta g,h) = AdU(g,\delta g)V(g,h) \\ &\operatorname{supp}\delta g \quad \text{outside} \quad \tilde{C}. \end{aligned} \tag{28}$$

  With this formula, the transition from the BPS–EG to the LQP net formalism has been achieved on the level of perturbation theory [23]. The algebraic content has been constructed in an auxiliary Fock space whose particle content is not necessarily identical with the physical particle content, and the adiabatic limit of the EG approach (which would have forced the coalescence of the two) has been avoided.
- *Perturbation as a deformation of free fields.* Having no control over the objects in the Bogoliubov axiomatics, we satisfy ourselves with existence and properties of causal power series for $S(g) := S(g,h)|_{h=0}$:

$$S(g) = \sum \frac{i^n}{n!} \int g(x_1)\ldots g(x_n) T W(x_1)\ldots W(x_n) \tag{29}$$

  which allows an iterative construction in n with W serving as the input. The main inductive step is the construction of the total diagonal part in $n+1$ order, assuming that the nth order time-ordered product has been fully (i.e. as an operator-valued distribution on all Schwarz test functions) constructed. Causality defines the $n+1$ order object on all test functions which vanish on a totally coalescent diagonal point [23]. The (Hahn–Banach) extension problem allows for totally locally supported terms with *a priori* undetermined coefficient. These local terms are often (as 'counter-terms') lumped together with the $n = 1$ term. Mere perturbative locality and unitarity requirements do not fix this ambiguity (i.e. perturbatively one always operators in Hilbert space†). Rather, the introduction of a suitable degree function allows us to control these ambiguities in terms of a finite number of physical parameters, at least in the case of so-called renormalizable interactions W with $\dim W \leqslant 4 = d$. Perturbation is a deformation around known theories which in the present

† This is not necessarily so in other (e.g. functional integral) formulations, where the connection with operator aspects of QT may get lost (even the introduction of cutoffs or regularizations is no assurance for maintaining it).

case are free fields. It only explores an infinitesimal neighbourhood around free fields and is not suited for deciding questions about mathematical existence. In fact, beyond deformation theory it is not physically compelling to implement the idea of interactions by coupling free fields to W in Fock space. Rather this is the perturbative way of introducing interactions and not a general consequence of the general framework. Indeed, the nonperturbative attempts based on modular theory use a different implementation of 'interaction', as will be shown later. The causal perturbation theory leads to the same renormalized correlation functions as, e.g., the one based on functional integrals. However, as shown in the following, the physical concepts and calculational rules are somewhat different. In particular, all differential identities (as equations of motion) can be used freely in the causal formulation, whereas this is not the case in the off-shell functional (Euclidean) approach. For the (m, s) free fields one may take any of the many possibilities in (8) independent of whether the field results from a classical Lagrangian (in which case its covariant transformation follows from the Euler equation of motions) or not. But since for given (m, s) there always exists a Lagrangian 'field coordinatization' in terms of which one may rewrite the given interaction W, one also does not lose anything if one starts from Lagrangians. The main benefit of the causal perturbation viewpoint lies in the fact that one liberates oneself from the moral obligation to repair something which came by quantization from classical theory. Instead the main question is how, by using the terms in the formal power series expansion, can one obtain something which is well defined in Fockspace, fulfils causality and unitarity requirements, and has the right to be called a time-ordered product of (the well-defined) W? This can be made more precise by saying it should coalesce with the naive time-ordered product of W if one smears them with test functions which have noncoalescent supports. So renormalization in the causal approach simply amounts to an (Hahn–Banach-like) extension of operator-valued distributions from the subspace of test functions with this restriction to all test functions. In addition one has to reparametrize the theory in terms of physical masses, charges and couplings, and use a field normalization which harmonizes with the asymptotic scattering interpretation or with the preservation of all selfinteractions in the adiabatic limit. Since there was no classical (bare) particle picture from quantization in this approach, there is also nothing to be repaired by dumping infinities. Hence the causal approach is finite in the distribution-theoretical sense, as is the Schwinger point-split methods, albeit much easier to handle than the latter. For dim $W \leqslant 4$ the procedure works in terms of obtaining a deformation theory with finitely many masses, charges and coupling parameters. To prove that this extension idea works in an inductive manner is not easy and the explanation of the necessary technical steps would throw this conceptually oriented presentation out of balance.

The above formal counting argument, if taken seriously as a definition of renormalizability, would rule out all massive higher spin $s \geqslant 1$ fields as candidates to be used for interaction polynomials W since there are no intertwiners from the Wigner particle to covariant local representations ψ with dim $\psi < 2$. For example, a massive $s = 1$ object in the vectormeson description has operator dimension dim $A_\mu = 2$ (the use of different intertwiners does not improve this increase of quantum versus classical dimension), so that any trilinear interaction involving A_μ (and lower spin) has dim $W \geqslant 5$. Fortunately, this barrier against renormalizability created by Wick polynomials of free fields involving $s \geqslant 1$ has an interesting loophole, namely it can be undermined by a 'cohomological trick' which consists of the following observation. One is asked to find a cohomological representation of, e.g., the

$(m, s = 1)$ physical Wigner space:

$$H_{Wigner} = \frac{\ker s}{\mathrm{i}ms} \qquad s^2 = 0. \tag{30}$$

Here s acts on H_{ext} and the Poincaré group is still covariantly represented on H_{ext} (the pseudo-unitary nature of the boost representors however turns out to be unavoidable). The transversality of the covariant inner product of the vectorpotential (which was the origin of $\dim A_\mu = 2$ instead of the classical dimension one) only emerges in the cohomological descent from H_{ext} to H_{Wigner}. The answer to the question why a *cohomological* extension and not another one (e.g. Gupta–Bleuler) which reduces the dimension to the classical value, lies in the hope that cohomological structures tend to be more stable under perturbative deformations. In other words, one expects a better chance to return to the physical space at the end of the perturbative calculations, in fact one expects the physical space to be the cohomology space. The simplest cohomological extension of the Wigner wavefunction space which allows a nilpotent operation† s with $s^2 = 0$, such that the physical transversality condition $p^\mu A_\mu(p) = 0$ follows from the application of s, needs, besides two scalar ghosts wavefunctions ω and $\bar{\omega}$, another scalar ghost field φ (often called the Stückelberg field):

$$\begin{aligned}
&(sA_\mu)(p) = p_\mu \omega(p) \\
&(s\omega)(p) = 0 \\
&(s\bar{\omega})(p) = p^\mu A_\mu(p) - \mathrm{i}m\varphi(p) \\
&(s\varphi)(p) = -\mathrm{i}m\omega(p).
\end{aligned} \tag{31}$$

One immediately realizes that $s^2 = 0$ and that $s(\cdot) = 0$ enforces the vanishing of ω and relates φ to $p^\mu A_\mu$. At this point there is no grading in the formalism, i.e. the ω and φ are simply ungraded wavefunctions. However, the functorial transition from Wigner theory to QFT in Fock space requires the introduction of a grading with $\deg \omega = 1$, $\deg \bar{\omega} = -1$, and $\deg A_\mu = 0$, with s transferring degree 1. The reason is that only with this grading assignment [27] does the s allow a natural tensor extension to multiparticle spaces with stable nilpotency,

$$s(a \otimes b) = sa \otimes b + (-1)^{\deg a} a \otimes sb \tag{32}$$

which insures the commutativity of the Wigner/Fock cohomological ascent and descent:

$$\begin{array}{ccc}
H_{ext} & \rightarrow & \mathcal{H}_{ext} \\
\downarrow & & \downarrow \\
H_{Wig} & \rightarrow & \mathcal{H}
\end{array} \tag{33}$$

where the calligraphic notation stands for the bosonic Fock space and its graded extension.

This prompts us to view the Fock space version δ of s as the image of a (pseudo)Weyl functor Γ as $\delta = \Gamma(s)$ and to write the δ in the spirit of a formal Noether symmetry charge Q:

$$Q = \int (\partial_\mu A^\mu(x) + m_a \phi(x)) \overset{\leftrightarrow}{\partial_0} \omega(x)\, \mathrm{d}^3x = Q^\dagger. \tag{34}$$

The experienced reader will easily recognize that we have arrived at a special version of the BRS formalism [24] which which remains unchanged by interactions [25].

The Fock space version of s yields an object δ of a differential algebra with $\delta^2 = 0$ which changes the Z-grading by one unit and acts on vectors and operators in H_{ext} similar to a global Noether charge:

$$\begin{aligned}
&\delta A = \mathrm{i}[Q, A] = \delta A \equiv \mathrm{i}\{QA - (-1)^{\deg A} AQ\} \\
&Q \quad \text{in} \quad H_{ext} \qquad Q^2 = 0.
\end{aligned} \tag{35}$$

† I apologise for using the letter s, in the following, for a special kind of cohomological (BRS) operation, after it served for the notation of spin as well as the (pre)Tomita operator; historical fidelity is sometimes a burden. The same fate will happen to δ later in this section.

Note that the nilpotency together with the formal hermiticity $Q = Q^\dagger$ prevents a positive inner product in $*$-representation of such algebras. It is customary (and helpful for mathematical control) to work with two inner products, one positive definite in order to stay with the mathematics of operators in Hilbert spaces, and a Krein operator η which is used to define another indefinite one as well as (pseudo)hermiticity. For many operators the two notions coalesce (they commute with Q), e.g. for all Poicaré generators except Lorentz boosts. In order to introduce interactions, one now uses the extended formalism in the same way as at the beginning of this section. For an interaction between vectormesons (for simplicity without additional matter) one may start with a trilinear expression (f_{abc} are independent couplings)

$$W^A = f_{abc} : A_{a\mu} A_{b\nu} \partial^\nu A_c^\mu \tag{36}$$

which in the extended space has dim $W = 4$. The important question to be answered now is: what is the criterion which selects the physical operators in $\mathcal{H}$ in every order of perturbation theory? Obviously they should commute with Q or rather the physical projection of the commutator should vanish. In addition to finding local operators with this property, one is interested in the S-matrix for the scattering of the massive particles which is the adiabatic limit of $S(g)$ for $g(x) \equiv \text{const} = g$. A sufficient condition on the operator-valued functional $S(g)$ which guarantees this property is that $S(g)$ commutes with Q up to surface terms in g which are localized in the collar (24). For the W and their time-ordered products which appear as integrands in these relations this means the validity of the following divergence equations:

$$\begin{aligned} &[Q, W(x)] = \mathrm{i}\partial_\mu^x W_1^\mu(x) \\ &[Q, T(W(x_1)\ldots W(x_n))] = \mathrm{i}\sum_{l=1}^{n} \partial_\mu^{x_l} T(W(x_1)\ldots W_1^\mu(x_l)\ldots W(x_n)). \end{aligned} \tag{37}$$

The W_1 must be constructed in the process of checking these relations. These equations were introduced by [26] and called ‘operator gauge invariance’. Although we will use these divergence relations, we will not follow this terminology because it creates the erroneous impression that a QFT involving massive vectormesons has to rely on a gauge principle in addition to renormalizability and the cohomological return to physics. It turns out that in contrast to what happens with low spin $s < 1$, the renormalization + cohomological descent requirement (the latter having no counterpart for low spin) are, in fact, so strongly restrictive, that not only the masses are forced to be equal and the couplings in (36) have to fulfil the Jacobi identity known from the Lie algebra structure, but all other couplings, including the quadrilinear couplings induced from the divergence equations, are such that as modulo renormalization terms they follow the pattern of classical gauge group theory, even though the group theory is not required by physical symmetries. However, the relation to the differential-geometric gauge structure is the opposite from that in the standard literature. Whereas classical gauge principles, which select among the many polynomial couplings (increasing number with increasing spin) involving vector fields those which nature (classical e.m.) prefers, usually enter QFT via quantization, the LQP approach produces a unique interaction between massive vectormesons in the way sketched before. In particular one obtains the inverse of the ’t Hooft renormalization statement namely the zero mass *(semi)classical limit of the unique perturbatively renormalizable massive vectormeson theory is a classical gauge theory*. Without going into more details [28] we will collect the important results of the above causal perturbation approach.

- The masses of the vectormesons are equal and the coupling among vectormesons and ghosts is determined by one coupling strength. The theory would show inconsistencies in higher than first order without the introduction of additional *physical* degrees of freedom.

The minimal (and perhaps only) possibility are (Higgs) scalars but without the usual vacuum expectations which are associated with 'Higgs' mechanism'.
- As expected from Schwinger's screening ideas [29], the physical $F_{\mu\nu}$-fields (those which commute with Q) have vanishing Maxwell charge and this would continue to be true in the presence of additional spinor matter.
- The uniqueness of the renormalizable spin $= 1$ part follows already from the specification of the physical particle content [28]; only the coupling between $s < 1$ matter introduces the usual additional parameters.

Comments. The results show that although the gauge point of view which requires the Higgs–Kibble mechanism ('fattening of photons by eating Goldstone bosons') is not incorrect, there is nothing physically intrinsic about it; it is a mnemotechnical device which allows a differential-geometrically inclined physicist rapid access to the perturbative results. It has the disadvantage that the necessity of the presence of additional physical degrees of freedom for reasons of consistency within renormalized perturbation theory (the Higgs degree of freedom) is not as convincing as in the present approach, in fact one usually puts the Higgs fields into the Lagrangian from the beginning. The present method leads to the same physical correlation functions but with a slightly different conceptual ring from the 'Mexican hat' arguments. The ghosts are clearly more recognizable as kinematical (via an extension of H_{Wigner}) auxiliary unphysical objects whereas the dynamical presence of additional physical degrees of freedom (the alias Higgs field, but without vacuum condensates) for matters of perturbative consistency becomes more manifest and the observable particle content receives greater emphasis. Classical differential-geometric concepts as the gauge idea are put into their proper place: they appear via Bohr's correspondence principle on the classical side as a result of the uniqueness of the implementation of perturbative renormalizability. Since gauge theories play a very prominent role, this point of view is not without interest. In fact it is close to the original viewpoint about massive vectormesons by Sakurai. The idea of the BRS like cohomological extension certainly takes care of those cases also covered by the gauge quantization and the Higgs–Kibble mechanism, but it may have a larger range of applicability to spin beyond one. The present method also suggests to consider the conceptually simpler (validity of scattering theory) massive case and approach the zero-mass situation with its infrared problems as a limiting case, i.e. the inverse of the Higgs approach. Since one knows that the physical charge-carrying fields in Maxwell-like theories have a noncompact spatial extension [13] (space-like cones with a semi-infinite string-like core), the physical massive fields cannot converge without the necessity of a prior modification. The attractive feature of such an idea is that such a modification becomes related to the decoupling of the Higgs particle.

There is a special feature of Abelian massive $s = 1$ theories with additional spinor matter which is absent in the non-Abelian case. Namely, in addition to the massive theory constructed in the analogous way with all matter fields being renormalizable, there also exists 'massive QED' for which the ψ-field cannot be simultaneously renormalizable (polynomially bounded correlation function with a dominating degree independent on perturbative order) and physical, i.e. commuting with Q. This massive QED has no Higgs degree of freedom which is apparantly necessary in order to have both properties.

A direct causal perturbative approach to $s = 1$ massless theories was recently formulated by Duetsch and Fredenhagen [22]. The necessity to avoid the (physically controversial) adiabatic limit requires the use of the full nonlinear BRS structure and to confront a situation in which (unlike as in the above case with bilinear Q) the position of the physical cohomology space keeps changing with the perturbative order. Lacking a fixed physical reference space (e.g. an incoming scattering space), the physical space only appears at the end as a representation space of a perturbative observable $*$-algebra. This construction was carried out in QED, but there is little doubt that with more work it also works for the non-Abelian case.

We do not, of course, claim that the BRS-like cohomological construction for the preservation of renormalizability in the face of higher spin advocated in these notes is less mysterious then the quantization gauge principle. It remains essentially magical why and how the cohomological trick produces local physical fields which at the end do not seem to be different from those obtained with the standard causal perturbation method, except that the latter cannot reconcile spin $= 1$ with renormalizability. However, despite its present magical touch, it is a bit closer to the spirit of LQP and perhaps less so to quantization and differential geometry. It keeps the attention on the unsolved infrared problems† and exposes the weird role of ghosts analogous to chemical catalyzers: they are introduced into the original physical problem in order to improve the W-powercounting and they are removed at the end without any visible trace [28].

This situation cries out for a deeper understanding without ghosts. From the more than 30-year struggle of physicists with this conceptual problem one should conclude that if there exists a formulation without ghosts in intermediate steps, then it cannot be anywhere near to the present formulation. In fact, the naturally ghost-free object is the S-matrix S which in contradistinction to the above transition operator of the causal approach $S(g)$ is on-shell. If one could find an iteration scheme directly for S which in intermediate steps avoids off-shell extrapolations, then this would be automatically ghost-free in every order. It would be a multivariable dispersion theoretical approach based on unitarity and crossing symmetry. The lowest order input consists of the on-shell tree diagrams (different from the off-shell W). Such an approach has only been carried out for $d = 1 + 1$ factorizing S-matrices where there exists a partial classification of admissible S-matrices even without the use of perturbation theory: the famous bootstrap-formfactor program of factorizable models. Outside of such restrictive situations a perturbative on-shell approach for S does not yet exist. The idea would be to use the perturbative ghost-free S-matrix in order to construct polarization-free generators of wedge algebras (PFGs). These are operators which are similar to free fields in that their one-time application onto the vacuum is a one-particle vector without admixtures of particle/antiparticle polarization clouds (see the last section). In the mentioned special case of factorizable models they are uniquely determined (see the last section) by the S-matrix via modular theory. Having generated the wedge algebras from the S-matrix, one can than use modular ideas to define and investigate a chiral conformal light-ray theory which is a canonical way associated with the wedge algebra. Although many of these statements sound futuristic, I think that this is the only way to avoid ghosts. One has to bypass the use of a Wick basis for the description of physical ghost-free operators as linear combinations of composite fields. Such a basis is not intrinsic and inevitably brings in the necessity of ghost field contributions. The approach dealing with algebras is the only basis-free intrinsic approach to the problem. The difficulty is the conversion of these rather abstract sounding ideas into a concrete computational scheme. The perturbative version of that only very incompletely understood on-shell scheme for low-spin renormalizable models which did not need ghosts in the old treatment should just reproduce the known renormalized results. Although our main present motivation for going to such extremes was to have a ghost-free renormalizable formalism for higher spin $s \geqslant 1$, the interest in it would by far exceed the present motivation. The idea of circumvention of the naive powercounting on W in terms of physical fields which rules out $s \geqslant 1$ by a radical reformulation of pertubation theory which directly leads to finite parametic physical theories for $s \geqslant 1$ is worth any effort

† From a physical point of view the aesthetic lure of differential geometry of fibre bundles in gauge theories is a bit dangerous, because it takes one away from the harder but physically more important infrared phenomena of the LQP of $s = 1$.

since it may turn out to be the tip of an iceberg. For further remarks we refer to [28].

We will return to this issue of generation of wedge algebras by modular methods in a more general context in the last section.

4. Modular origin of geometric and hidden symmetries

From the wedge localization in section 2 we have seen that the modular objects associated to a standard (cyclic and separating vector Ω) pair $(\mathcal{A}(\mathcal{O}), \Omega)$ has, under certain circumstances, a geometrical significance, e.g. for the wedge in a massive (Poincaré-invariant) theory, or the double cone in a massless (conformally invariant) theory. This suggests the question whether all space-time symmetries (diffeomorphisms) can be viewed as having a modular algebraic origin, i.e. if they can be thought of as originating from the relative positions of individual algebras in a net. *This would elevate space-time from its role of merely indexing individual algebras in the net, to a structure which is on the one hand more intimately related with the physical aspects of LQP, and on the other hand presents already structural properties whose understanding seems to be a prerequisite for the formulation of the elusive 'quantum gravity'*. It turns out that in chiral conformal theories the Moebius group, together with the net on which it acts, can be constructed from only two properly positioned algebras which give rise to two 'half-sided modular inclusions' (see below). In fact, mathematically, the world of chiral conformal nets is equivalent with the classification of all 'standard half-sided modular inclusions'. In this conformal setting the Haag duality is automatic and there is no spontaneous symmetry breaking. The analogue in the higher-dimensional case is to assume wedge duality (always achievable, as previously mentioned, by maximization) and to prove the equality of the modular group with the Lorentz boost without assuming (as Bisognano and Wichmann did) that the algebras are generated by local fields. Presently this cannot be done without making additional assumptions, i.e. assumptions which cannot be expressed in terms of modular positions only, but are suggested by space-time geometry [55]. Amazingly one again succeeds in building up the whole Poincaré group as well as the net from a small finite number of algebras in appropriate modular positions (using modular inclusions and modular intersections).

Since modular groups exist for each space-time region one may ask about their physical interpretation. Let us start by posing the opposite question in a context where there are geometric candidates without obvious modular origin. In chiral conformal theories one has a rich supply of diffeomorphisms of the circle which have been around since the beginning of the 1970s. The way these mathematical structures were discovered by physicists is somewhat bizarre and confusing. It is interesting to take a brief look at the history by permitting a short interlude, before presenting our modular interpretation.

Apart from some early work of mathematicians (Gelfand, Fuchs) on diffeomorphisms of S^1 and their associated Witt algebra (infinitesimal diffeomorphisms without the central extension), the first observation by a physicist of this Witt algebra structure was made in the Veneziano dual S-matrix model by Virasoro [30]. At that time it was realized that the on-shell dual S-matrix model allowed for a nice off-shell presentation in terms of a massless free-field theory in $d = 1 + 1$. Parallel to this, but without any interrelation, there were detailed field theoretic investigations of the representation of conformal generators in terms of the energy momentum tensor T and their action e.g. on the Thirring fields [33] and the problem (formulated in Lowenstein's thesis and going back to Greenberg) of classifying so-called 'Lie fields' [34], the predecessors of what in the rediscovered version 25 years later were called W-fields. The next contribution came again from the dual model calculations and consisted in the correct computation of the central term (for free massless fermions) which was previously overlooked [35]. My own contribution was the computation in 1973 of the general structure of the T–T commutation relation in chiral conformal theories as a structural consequence of translational covariance and causality which I presented together with other results at the January 1974, V Brazilian Symposium in Rio de Janeiro [36]. Apart from not knowing the aforementioned free-fermion results, my motivation was quite different and consisted in the search for a nontrivial 'Lie field' of which the energy momentum tensor was the first illustration†. In the same year the conformal block decomposition was discovered (called the decomposition of local fields into nonlocal components) which solved the Einstein 'causality paradox in conformal quantum field theory' by noticing [37] that local fields were irreducible only with respect to a finite neighbourhood of the identiy but not with respect to the centre of the covering of $SL(2,R) \times SL(2,R)$. The illustration of this decomposition theory by nontrivial models (minimal models) beyond exponential Bose fields had to wait another ten years [38]. By that time the increased knowledge by physicists about infinite-dimensional Lie algebras (affine algebras, diffeomorphism algebras) was leaving its mark on low-dimensional QFT. This also had, besides many gains, one disadvantage because the use of those infinite-dimensional Lie algebras separated these low-dimensional QFT sharply from a higher-dimensional standard type of QFT to which such structures are not available. The modular point of view which I will present in the following admits a higher-dimensional analogue and incorporates conformal and factorizing theories back into the framework of general QFT.

Returning to the modular issue, let us look at a special subgroup whose Lie algebra is isomorphic to that of the Moebius group. Its action on the circle is

$$z \to \sqrt{\frac{a + bz^2}{c + dz^2}} \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SU(1,1) \tag{38}$$

where the cuts connecting both poles and zeros are chosen outside the unit circle. In fact this defines a two-fold covering of the Moebius group. Given an interval, its square root (inverse image of $z \to z^2$) consists of two disjoint intervals which are separately left invariant under the above transformation group. The obvious conjecture is of course that (as for the case of a single interval) the covering dilation subgroup is the modular group of the pair $(\mathcal{A}(I_1 \cup I_2), \Omega)$. But this cannot be, because this action restricted to one interval is the same as that of the dilation in the Moebius group but this, according to a theorem by Takesaki [44] is not possible if the vacuum state fulfils the Reeh–Schlieder property of being cyclic and separating for only one interval. Since it never happens that two disjoint square-root intervals are contained in one interval of another such pair, there will be no contradiction with the lack of the Reeh–Schlieder property for one interval. A (quasifree) state on the Weyl algebra (which we take as an illustration of a simple conformal model) which is invariant under the above covering transformation [32] is easily found in terms its two-point function which belongs to the following scalar product:

$$\langle f, g \rangle = \int \frac{f(x)g(y)}{[(x-y)(1+xy)+\mathrm{i}\varepsilon]^2}(1+x^2)(1+y^2)\,\mathrm{d}x\,\mathrm{d}y \tag{39}$$

where we used the linear presentation instead of the circular one ($SL(2,R)$ instead of $SU(1,1)$). This is to be compared with the standard inner product belonging to the vacuum representation

$$\langle f, g \rangle_0 = \int \frac{f(x)g(y)}{(x-y+\mathrm{i}0)^2}\,\mathrm{d}x\,\mathrm{d}y. \tag{40}$$

† The reason why many field theoretical results on low-dimensional field theories were only published in conference proceedings was sociological and not scientific. Low-dimensional field theory for the benefit of higher-dimensional S-matrix models was considered of greater physical relevance than its use as a theoretical laboratory for the test of general ideas on interactions, a point of view which was later uphold by string theorists.

One easily checks that this inner product belongs to the same symplectic form as the standard one, namely

$$\omega(f,g) = \mathrm{Im}\,\langle f,g\rangle = \int fg'\mathrm{d}x = \omega_0(f,g) = \mathrm{Im}\,\langle f,g\rangle_0. \tag{41}$$

As for the standard case the criterion for a Fock representation is that the inner product can be represented in terms of ω with the help of a complex structure I_0, $I_0^2 = -1$, with

$$\begin{aligned}&\langle f,g\rangle_0 = \omega(I_0 f,g) = -\omega(f,I_0 g)\\&(I_0 f)(x) \equiv \int \frac{-1}{(x-y+\mathrm{i}\varepsilon)} f(y)\,\mathrm{d}y\end{aligned} \tag{42}$$

the analogous statement holds for $\langle f,g\rangle$ with I_0 replaced by I

$$\begin{aligned}&I = \Gamma^{-1}\circ I_0\circ\Gamma\\&(\Gamma f)(x) \equiv \int f\left(\frac{x}{2} + \mathrm{sign}(x)\sqrt{1+\left(\frac{x}{2}\right)^2}\right)\\&\langle\Gamma f,\Gamma g\rangle_0 = \langle f,g\rangle.\end{aligned} \tag{43}$$

The changed inner product defines a changed quasifree state on the Weyl algebra. The proof that the covering dilation

$$\begin{aligned}&U(\lambda) = \Gamma^{-1}\circ V(\lambda)\circ\Gamma\\&(V(\lambda)f)(x) = f(\lambda x)\end{aligned} \tag{44}$$

is indeed the modular group for the algebra of the disjoint intervals $[-\infty,-1]\cup[0,1]$ in this quasifree state, we only have to check the appropriate KMS condition. From:

$$\lim_{\theta\uparrow 2\pi}\langle U(\lambda)f,g\rangle_1 = \lim_{\theta\uparrow 2\pi}\langle V(\lambda)\circ\Gamma(f),\Gamma(g)\rangle = \langle\Gamma(g),\Gamma(f)\rangle = \langle g,f\rangle_1 \tag{45}$$

one sees that the $U(\lambda)$ fulfils the KMS condition if both f and g are from one of the two intervals since Γ transforms the space of $[0,1]$ localized functions into $[-\infty,-1]$ localized ones and vice versa.

This situation is very interesting, since although the chiral diffeomorphisms allows no geometric generalization to diffeomorphisms in higher dimensional LQP, the disconnected (and multiply connected) algebras have modular groups which act in a nonpoint-like manner inside these disconnected local regions†. This is what we mean by 'hidden symmetries'. There is another closely related aspect which strengthens the physical relevance of disconnected regions. It was well known for some time [1] that such situations break Haag duality i.e.

$$\mathcal{A}((I_1\cup I_2)') \subset \mathcal{A}(I_1\cup I_2)'$$

is a genuine inclusion if the net $\mathcal{A}$ has nontrivial superselection rules. For models resulting from the maximal extension of the Abelian current algebra the mechanism which causes this obstruction against Haag duality has been completely analysed in [46]. Very recently this has been understood in complete generality (for rational theories, i.e. those with a finite number of sectors) in [45] by using very powerful methods of subfactor theory. In the context of the above use of 'geometric states', one would conjecture that their lack of cyclicity leads to a Jones projector which contains the information about the additional superselection sectors, but this remains to be seen.

In the following we will look at two more illustrations of modular constructions.

† Observable algebras in disconnected regions have also played a role as indicators of the presence of charge sectors [31, 43, 46].

As a reference wedge we may take the wedge $W(l_1, l_2)$ spanned by the light-like vectors $l_{1,2} = e_\pm = (1, 0, 0, \pm 1)$, in which case we call z, t the longitudinal and x, y the transversal coordinates (the light-like characterization of wedges is convenient for the following). This situation suggests to decompose the Poincaré group generators into longitudinal, transversal and mixed generators

$$P_\pm = \frac{1}{\sqrt{2}}(P_0 \pm P_z) \qquad M_{0z};\, M_{12},\, P_i;\, G_i^{(\pm)} \equiv \frac{1}{\sqrt{2}}(M_{i0} \pm M_{iz}) \qquad i = 1, 2. \tag{46}$$

The generators $G_i^{(\pm)}$ are precisely the 'translational' pieces of the Euclidean stability groups $E^{(\pm)}(2)$ of the two light vectors $e_\pm$ which appeared in Wigner's representation theory for zero-mass particles. More recently, these 'translations' inside the homogenous Lorentz group appeared in the structural analysis of 'modular intersections' of two wedges [39, 40]. Apart from the absence of the positive spectrum condition, its role is analogous to that of the true translations $P_\pm$ with respect to half-sided 'modular inclusions' [40].

As one reads off from the commutation relations, P_i, $G_i^{(+)}$, $P_\pm$ have the interpretation of a central extension of a transversal 'Galilei group'† with the two 'translations' $G_i^{(+)}$ representing the Galilei generators, P_+ the central 'mass' and P_- the 'nonrelativistic Hamiltonian'. The longitudinal boost M_{0z} scales the Galilei generators $G_i^{(+)}$ and the 'mass' P_+. Geometrically the $G_i^{(+)}$ change the standard wedge (it tilts the logitudinal plane) and the corresponding finite transformations generate a family of wedges whose envelope is the half-space $x_- \geqslant 0$. The Galilei group together with the boost M_{0z} generate an eight-parametric subgroup $G^{(+)}(8)$ inside the ten-parametric Poincaré group‡:

$$G^{(+)}(8):\ P_\pm,\ M_{0z};\ M_{12},\ P_i;\ G_i^{(+)}. \tag{47}$$

The modular reflection J transforms this group into an isomorphic $G^{(-)}(8)$.

The Galilean group is usually introduced as a 'contraction' of the Poincaré group. But as the present discussion, the wedge (or rather as in the following remarks, two wedges in a special modular intersection position) shows, it also appears as a genuine subgroup of the Poincaré group. The latter fact seems to be less known.

All observations have interesting generalizations to the conformal group in massless theories in which case the associated natural space-time region is the double cone.

This subgroup $G^{(+)}(8)$ is intimately related to the notion of modular intersection see [39, 40]. Let l_1, l_2 and l_3 be three linear independent light-like vectors and consider two wedges $W(l_1, l_2)$, $W(l_1, l_3)$ with Λ_{12} and Λ_{13} the associated Lorentz boosts. As a result of this common l_1 the algebras $\mathcal{N} = \mathcal{A}(W(l_1, l_2))$, $\mathcal{M} = \mathcal{A}(W(l_1, l_3))$ have a modular intersection with respect to the vector Ω. Then $(\mathcal{N} \cap \mathcal{M}) \subset \mathcal{M}, \Omega)$ is a so-called modular inclusion [40,41]. Identifying $W(l_1, l_2)$ with the above standard wedge, we notice that the longitudinal generators $P_\pm$, M_{0z} are related to the inclusion of the standard wedge algebra into the full algebra $\mathcal{B}(H)$, whereas the Galilei generators $G_i^{(+)}$ are the 'translational' part of the stability group of the common light vector l_1 (i.e. of the Wigner light-like little group).

To simplify the situation let us take $d = 1 + 2$ with $\mathcal{G}(4)$, in which case there is only one Galilei generator G. In addition to the 'visible' geometric subgroup of the Poincaré group, the modular theory produces a 'hidden' symmetry transformation $U_{\mathcal{N}\cap\mathcal{M},\mathcal{M}}(a)$ which belongs to

† G are only Galilean in the transverse sense; they tilt the wedge so that one of the light-like directions is maintained but the longitudinal plane changes.

‡ The Galilean group is usually introduced as a 'contraction' of the Poincaré group. But as the present discussion about the wedge (or rather the following remarks about two wedges in a special modular intersection position) shows, it also appears as a genuine subgroup of the Poincaré group. The latter fact seems to be less known.

a region which is an intersection of two wedges:

$$U_{\mathcal{N}\cap\mathcal{M},\mathcal{M}}(a) := \exp\left(\frac{\mathrm{i}a}{2\pi}(\ln\Delta_{N\cap M} - \ln\Delta_M)\right) \tag{48}$$

is a unitary group with positive generator. Moreover one has:

$$U_{\mathcal{N}\frown\mathcal{M},\mathcal{M}}(1-\mathrm{e}^{-2\pi t}) = \Delta_{\mathcal{M}}^{it}\Delta_{\mathcal{N}\frown\mathcal{M}}^{-it} \tag{49}$$

$$U_{\mathcal{N}\cap\mathcal{M},\mathcal{M}}(\mathrm{e}^{-2\pi t}a) = \Delta_{\mathcal{M}}^{it}U_{\mathcal{N}\cap\mathcal{M},\mathcal{M}}(a)\Delta_{\mathcal{M}}^{-it} \tag{50}$$

$$\mathrm{Ad}\,U_{\mathcal{N}\cap\mathcal{M},\mathcal{M}}(-1)\,(\mathcal{M}) = \mathcal{N}\cap\mathcal{M} \tag{51}$$

and

$$J_{\mathcal{M}}U_{\mathcal{N}\frown\mathcal{M},\mathcal{M}}(a)J_{\mathcal{M}} = U_{\mathcal{N}\frown\mathcal{M},\mathcal{M}}(-a). \tag{52}$$

Similar results hold for $\mathcal{N}$ replacing $\mathcal{M}$. Due to the intersection property we finally have the commutation relation

$$[U_{\mathcal{N}\frown\mathcal{M},\mathcal{M}}(a), U_{\mathcal{N}\frown\mathcal{M},\mathcal{N}}(b)] = 0 \tag{53}$$

which enables one to define the unitary group

$$U_{\mathcal{N}\frown\mathcal{M}}(a) = U_{\mathcal{N}\frown\mathcal{M},\mathcal{M}}(-a)U_{\mathcal{N}\frown\mathcal{M},\mathcal{N}}(a). \tag{54}$$

This latter group can be rewritten as

$$U_{\mathcal{N}\frown\mathcal{M}}(1-\mathrm{e}^{-2\pi t}) = \Delta_{\mathcal{M}}^{it}\Delta_{\mathcal{N}}^{-it} \tag{55}$$

and thereby recognized to be in our physical application the one-parameter Galilean subgroup G (47) in the above remarks.

Now we notice that for $a < 0$

$$\begin{aligned}\mathrm{Ad}\,U_{\mathcal{N}\frown\mathcal{M},\mathcal{M}}(a)(\mathcal{M}) &= \mathrm{Ad}\,\Delta_{\mathcal{M}}^{-\mathrm{i}(\frac{1}{2\pi}\ln -a)}U_{\mathcal{N}\frown\mathcal{M},\mathcal{M}}(-1)(\mathcal{M}) \\ &= \mathrm{Ad}\,\Delta_{\mathcal{M}}^{-\mathrm{i}(\frac{1}{2\pi}\ln -a)}(\mathcal{N}\cap\mathcal{M}).\end{aligned} \tag{56}$$

Because $\Delta_{\mathcal{M}}^{it}$ acts geometrically as Lorentz boosts, we have full knowledge of the geometrical action of $U_{\mathcal{N}\frown\mathcal{M},\mathcal{M}}(a)$ on $\mathcal{M}$ for $a < 0$. For $a > 0$ we notice

$$\begin{aligned}\mathrm{Ad}\,U_{\mathcal{N}\frown\mathcal{M},\mathcal{M}}(1)(\mathcal{M}) &= \mathrm{Ad}\,U_{\mathcal{N}\frown\mathcal{M},\mathcal{M}}(2)(\mathcal{M}\cap\mathcal{N}) = \mathrm{Ad}\,J_{\mathcal{M}}J_{\mathcal{N}\frown\mathcal{M}}(\mathcal{M}\cap\mathcal{N}) \\ &= \mathrm{Ad}\,J_{\mathcal{M}}(\mathcal{M}'\cup\mathcal{N}')\end{aligned} \tag{57}$$

and again, due to the geometrical action of J_M we have a geometrical action on $\mathcal{M}$ for $a > 0$.

$$\mathrm{Ad}\,U_{\mathcal{N}\cap\mathcal{M},\mathcal{M}}(a)(\mathcal{M}) = \mathrm{Ad}\,\Delta_{\mathcal{M}}^{-\mathrm{i}(\frac{1}{2\pi}\ln a)}J_{\mathcal{M}}(\mathcal{M}'\cup\mathcal{N}'). \tag{58}$$

From these observations and with $U_{\mathcal{N}\cap\mathcal{M},\mathcal{M}}(1-\mathrm{e}^{-2\pi t}) = \Delta_{\mathcal{M}}^{it}\Delta_{\mathcal{M}\cap\mathcal{N}}^{-it}$ we get for $t < 0$:

$$\mathrm{Ad}\,\Delta_{\mathcal{N}\cap\mathcal{M}}^{it}(\mathcal{M}) = \mathrm{Ad}\,\Delta_{\mathcal{M}}^{(-\frac{\mathrm{i}}{2\pi}\ln(\mathrm{e}^{-2\pi t}-1))}J_{\mathcal{M}}(\mathcal{M}'\cup\mathcal{N}') \tag{59}$$

and in the case of $t > 0$:

$$\mathrm{Ad}\,\Delta_{\mathcal{N}\cap\mathcal{M}}^{it}(\mathcal{M}) = \mathrm{Ad}\,\Delta_{\mathcal{M}}^{(-\frac{\mathrm{i}}{2\pi}\ln(1-\mathrm{e}^{-2\pi t}))}(\mathcal{N}\cap\mathcal{M}). \tag{60}$$

Similar results hold for $\mathcal{N}$ replacing $\mathcal{M}$. With the same methods we get:

$$\begin{aligned}\mathrm{Ad}\,\Delta_{\mathcal{N}\cap\mathcal{M}}^{it}\Delta_{\mathcal{N}}^{is}(\mathcal{M}) &= \mathrm{Ad}\,\Delta_{\mathcal{N}\cap\mathcal{M}}^{it}\Delta_{\mathcal{N}}^{is}\Delta_{\mathcal{M}}^{-is}(\mathcal{M}) \\ &= \mathrm{Ad}\,\Delta_{\mathcal{N}\cap\mathcal{M}}^{it}U_{\mathcal{M}\cap\mathcal{N}}(\mathrm{e}^{-2\pi s}-1)(\mathcal{M})\end{aligned} \tag{61}$$

where $U_{\mathcal{N}\cap\mathcal{M}}$ is the one-parameter Lorentz subgroup (the Galilei subgroup G in (47) associated with the modular intersection. This gives:

$$\begin{aligned}\mathrm{Ad}\,\Delta_{\mathcal{N}\cap\mathcal{M}}^{it}\Delta_{N}^{is}(\mathcal{M}) &= \mathrm{Ad}\,U_{\mathcal{M}\cap\mathcal{N}}(\mathrm{e}^{-2\pi t}(\mathrm{e}^{-2\pi s}-1))\Delta_{\mathcal{N}\cap\mathcal{M}}^{it}(\mathcal{M}) \\ &= \mathrm{Ad}\,U_{\mathcal{M}\cap\mathcal{N}}(\mathrm{e}^{-2\pi t}(\mathrm{e}^{-2\pi s}-1))\Delta_{\mathcal{M}}^{-\frac{1}{2\pi}\ln(1-\mathrm{e}^{-2\pi t})}(\mathcal{M}\cap\mathcal{N})\end{aligned} \tag{62}$$

if $t > 0$ and similar for $t < 0$.Therefore we get a geometrical action of $\Delta^{it}_{\mathcal{N}\cap\mathcal{M}}$ on $\mathrm{Ad}\,\Delta^{is}_{\mathcal{N}}(\mathcal{M})$.

A look at the proof shows that the essential ingredients are the special commutation relations. Due to

$$\Delta^{it}_{\mathcal{M}\cap\mathcal{N}} = \Delta^{it}_{\mathcal{M}} U_{\mathcal{N}\cap\mathcal{M},\mathcal{M}}(1 - \mathrm{e}^{-2\pi t}) = \Delta^{it}_{\mathcal{M}} J_{\mathcal{M}} U_{\mathcal{N}\cap\mathcal{M},\mathcal{M}}(\mathrm{e}^{-2\pi t} - 1) J_{\mathcal{M}} \quad (63)$$

and the well established geometrical action of $\Delta^{it}_{\mathcal{M}}$ and $J_{\mathcal{M}}$, it is enough to consider the action of $U_{\mathcal{N}\cap\mathcal{M},\mathcal{M}}$ or similarly $U_{\mathcal{N}\cap\mathcal{M},\mathcal{N}}$. For these groups we easily get

$$\mathrm{Ad}\,U_{\mathcal{N}\cap\mathcal{M},\mathcal{M}}(a)\Delta^{is}_{\mathcal{N}}\Delta^{-it}_{\mathcal{M}}(\mathcal{N}) = \mathrm{Ad}\,\Delta^{is}_{\mathcal{N}}\Delta^{it}_{\mathcal{M}} U_{\mathcal{N}\cap\mathcal{M},\mathcal{M}}(\mathrm{e}^{-2\pi(s+t)}a)(\mathcal{N}) \quad (64)$$

and due to the above remarks the geometrical action of $\Delta^{it}_{\mathcal{N}\cap\mathcal{M}}$ on the algebras of the type $\mathrm{Ad}\,\Delta^{is}_{\mathcal{N}}\Delta^{-it}_{\mathcal{M}}(\mathcal{M})$.

Now, the light-like translations $U_{transl_1}(a)$ in l_1 direction fulfil the positive spectrum condition and map $\mathcal{N}\cap\mathcal{M}$ into itself for $a > 0$. Therefore, we have the Borchers commutator relations with $\Delta^{it}_{\mathcal{M}\cap\mathcal{N}}$ and get

$$\mathrm{Ad}\,\Delta^{it}_{\mathcal{N}\frown\mathcal{M}} U_{transl_1}(a)(\mathcal{M}) = \mathrm{Ad}\,U_{transl_1}(\mathrm{e}^{-2\pi t}a)\Delta^{it}_{\mathcal{N}\frown\mathcal{M}}(\mathcal{M}). \quad (65)$$

The additivity of the net tells us that taking unions of the algebra corresponds to the causal unions of localization regions. The assumed duality allows us to pass to causal complements and thereby to intersections of the underlying localization regions. Therefore the algebraic properties above transfer to unions, causal complements and intersections of regions. We finally get [32] the following theorem.

Theorem 1. *Let $\mathcal{R}$ be the set of regions in $\boldsymbol{R}^{1,2}$ containing the wedges $W[l_1, l_2]$, $W[l_1, l_3]$ and which is closed under:*

(a) Lorentz boosting with $\Lambda_{12}(t)$, $\Lambda_{13}(s)$,
(b) intersection,
(c) (causal) union,
(d) translation in l_1 direction,
(e) causal complement.

Then $\Delta^{it}_{W[l_1,l_2]\cap W[l_1,l_3]}$ maps sets in $\mathcal{R}$ onto sets in $\mathcal{R}$ in a well-computable way and extends the subgroup (47) by a 'hidden symmetry'.

Similarly we can look at a $(1+3)$-dim quantum field theory. Then we get the same results as above for the modular theory to the region $W[l_1, l_2] \cap W[l_1, l_3] \cap W[l_1, l_4]$, where l_i are four linear independent light-like vectors in $\boldsymbol{R}^{1,3}$. Moreover in this case the set $\mathcal{R}$ contain $W[l_1, l_2]$, $W[l_1, l_3]$ and $W[l_1, l_4]$ is closed under boosting with $\Lambda_{12}(t)$, $\Lambda_{13}(s)$, $\Lambda_{14}(r)$.

The arguments are based on the Borchers commutation relation and modular intersection theory and also apply if we replace modular intersection by modular inclusion. One easily recovers in this way the results of Borchers and Yngvason [42], who found an illustration of hidden symmetries in thermal chiral conformal QFT. (Note that in thermal situations we have no simple geometrical interpretation for the commutants as the algebra to causal complements. Therefore, in these cases, we have to drop (e) in the above theorem.)

The final upshot of this section is to show that there might be a well-defined meaning of a geometrical action of modular groups by restricting on certain subsystems.

For conformal LQP in any dimension, one obtains a generalization of the previous situation. In particular, the modular group with respect to the vacuum of the double-cone algebra is geometric [1]. Consider now a double-cone algebra $\mathcal{A}(\mathcal{O})$ generated by a free massless field (for $s = 0$ take the infrared convergent derivative). Then according to the previous remark, the modular objects of $(\mathcal{A}(\mathcal{O}), \Omega)_{m=0}$ are well known. In particular, the modular group is a one-parametric subgroup of the proper conformal group. The massive double-cone algebra together with the (wrong) massless vacuum has the same modular group, σ_t, however its action on smaller massive subalgebras inside the original one is not describable in terms of the previous subgroup. In fact, the geometrical aspect of the action is wrecked by the breakdown of Huygens principle, which leads to a nonlocal reshuffling inside $\mathcal{O}$ but still is local in the sense of keeping the inside and its causal complement apart. This mechanism can be shown to lead to a pseudo-differential operator for the infinitesimal generator of σ_t whose highest term still agrees with conformal zero-mass differential operator. We are, however, interested in the modular group of $(\mathcal{A}(\mathcal{O}), \Omega)_m$ with the massive vacuum which is different from the that of the wrong vacuum by a Connes cocycle. We believe that this modular cocycle will not wreck the pseudo-differential nature and that as a consequence the geometric nature of the conformal situation will still be asymptotically true near the horizon of the double cone, however, we are presently not able to show this. This modular aspect of the horizon could be linked with what people think should be the quantum version of the Bekenstein–Hawking classical entropy considerations, in particular the ideas about 'holographic properties'. To be more precise, we expect that even for double cones in Minkowski space (i.e. without a classical Killing vector as for black holes) there will be a finite relative quantum entropy as long as one allows for a 'collar' between the double cone and its space-like complement and that with vanishing size of this collar these entropies will diverge in such a way that ratios (e.g. for differently sized double cones) will stay finite and be determined by the conformal limits. In this way one could hope to prove that, e.g., the speculations about entropy, holography and the occurrence of the central terms in the energy momentum commutation relations are nonperturbative generic properties of ordinary LQP [6]. For the thermal aspects this is, of course, well known.

The modular group structure also promises to clarify some points concerning the physics of the Wightman domain properties [10]. In fact these groups act linearly on the 'field space' i.e. the space generated by applying a local field on the vacuum. Therefore this space, which is highly reducible under the Poincaré group, may (according to a conjecture of Fredenhagen, based on the results in [47]) in fact carry an irreducible representation of the union of all modular groups (an infinite-dimensional group $\mathcal{G}_{mod}$ which contains in particular all local space-time symmetries). The equivalence of fields with carriers of irreducible representations of an universal $\mathcal{G}_{mod}$ would add a significant conceptual element to LQP and give the notion of quantum fields a deep role which goes much beyond that of being simply generators of local algebras. Our arguments suggest that in chiral conformal QFT $\mathcal{G}_{mod}$ includes all local diffeomorphism.

A related group theoretical approach to LQP which uses both modular groups and modular involutions in order to formulate a new selection principle for states ('the condition of geometric modular action') was proposed in [48]. In addition to the modular groups which leave the defining local algebras invariant, these authors obtain a discrete group (from the conjugations) which transform the (space-time) index set. All these true QFT properties remain invisible in any quantization approach. Combining modular theory with scattering theory, the actual J together with the incoming J^{in} can be used to obtain a new framework for nonperturbative interactions [10]. This last topic will be presented in the following section; more details can be found in a separate paper together with H-W Wiesbrock [52].

5. Constructive modular approach to interactions

The starting observation for relating the modular structure of LQP nets to interactions is that the latter is solely contained in those anti-unitary reflections of the full Poincaré group which contain the time reversal. The continuous part (as well as those reflections which do not involve time) is, thanks to the fact that scattering (Haag–Ruelle, LSZ) theory is a consequence of LQP, the same for the free incoming particles as for the interacting net [46]:

$$U(\Lambda, a) = U(\Lambda, a)^{in} \tag{66}$$

$$\begin{aligned} J &= S_{sc} J^{in} \\ S_T &= J \Delta^{\frac{1}{2}} \\ S_T A\Omega &= A^* \Omega. \end{aligned} \tag{67}$$

Here S is the scattering matrix. The subscript T is used in order to distinguish the Tomita operator from the scattering matrix and the J is the Tomita reflection for interacting wedge algebras whereas J^{in} refers to the algebra generated by the incoming free field. The standard point of view, where the interaction is introduced in terms of a pair of Hamiltonians (Lagrangians) H, H_0, accounts for the interaction in another (more perturbative) way which uses different states. It is well known that this standard perturbative approach cannot be directly formulated in infinite space because translational invariance together with invariance of the vacuum is in contradiction with the existence of another Hamiltonian H once a bilinear H_0 has been specified (Haag's theorem). In perturbation theory this is not a serious obstacle; it is formally taken care of by leaving out the pure vacuum Feynman graphs or more carefully by using the Feynman–Gell-Mann formula in a quantization box and taking the thermodynamic limit. The modular approach does not have this problem.

The most promising candidates for a modular construction are obviously massive theories with a known S-matrix, i.e. models which permit a bootstrap construction of S on its own, without using the off-shell fields or local operators. For such S-matrix integrable models, there already exists a constructive formfactor program which goes back to Karowski and Weisz and has been significantly extended by Smirnov [53, 54]. It uses suggestive prescriptions and assumptions within the dispersion theoretical LSZ framework.

Since the bulk of the LSZ formalism is a consequence of the more basic algebraic QFT, it is reasonable to ask if our modular localization framework is capable of shedding additional light on this programme, in particular whether it can be understood as a special (analytically simple) case of a more general nonperturbative construction without the restriction to $d = 1+1$ factorizing theories [10]. The crucial vehicle which carries the off-shell modular and thermal properties of wedge regions to on-shell crossing properties of formfactors are very subtle polarization-free wedge generators (PFG) which we will now explain.

Let us start with a very simple-minded generalization of free fields in $d = 1 + 1$. For the latter we use the notation:

$$\begin{aligned} A(x) &= \frac{1}{\sqrt{2\pi}} \int (\mathrm{e}^{-\mathrm{i}px} a(p) + \text{h.a.}) \frac{\mathrm{d}p}{2\omega} \\ &= \frac{1}{\sqrt{2\pi}} \int (\mathrm{e}^{-\mathrm{i}m\rho sh(\chi-\theta)} a(\theta) + \text{h.a.})\, \mathrm{d}\theta \qquad x^2 < 0 \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathsf{C}} \mathrm{e}^{-\mathrm{i}m\rho sh(\chi-\theta)} a(\theta)\, \mathrm{d}\theta \qquad \mathsf{C} = \mathbb{R} \cup \{-\mathrm{i}\pi + \mathbb{R}\} \end{aligned} \tag{68}$$

where in the second line we have introduced the x- and momentum-space rapidities and specialized to the case of space-like x, and in the third line we used the analytic properties of the exponential factors in order to arrive at a compact and (as it will turn out) useful contour representation. Note that the analytic continuation refers to the c-number function, whereas the formula $a(\theta - \mathrm{i}\pi) \equiv a^*(\theta)$ is a definition and has nothing to do with analytic continuations of operators†.

With this notational matter out of the way, we now write down our ansatz

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathrm{C}} \mathrm{e}^{-\mathrm{i}m\rho sh(\chi-\theta)} Z(\theta)\, \mathrm{d}\theta \tag{69}$$

$$\begin{aligned} &Z(\theta)\Omega = 0 \qquad Z(\theta_1)Z(\theta_2) = S_{Z,Z}(\theta_1-\theta_2)Z(\theta_2)Z(\theta_1) \\ &Z(\theta_1)Z^*(\theta_2) = \delta(\theta_1-\theta_2) + S_{Z,Z^*}(\theta_1-\theta_2)Z^*(\theta_2)Z(\theta_1). \end{aligned} \tag{70}$$

For the moment the S are simply Lorentz-covariant (only rapidity differences appear) functions which for algebraic consistency fulfil unitarity $\overline{S(\theta)} = S(-\theta)$. We assume (for simplicity) that the state space contains only one type of particle.

A field operator $F(x)$ is called 'one-particle polarization free' (PF) if $F(x)\Omega$ and $F^*(x)\Omega$ have only one-particle components (for any one of the irreducible particle spaces in the theory).

Obviously the above $F(x)$ with $Z(\theta)\Omega = 0$ (but yet without the algebraic relations which specialize the interactions to the relativistic counterpart of quantum mechanical pair interactions) is the most general PF in $d = 1 + 1$. The PF property is an on-shell concept, but note that nothing is required about the nature of state vectors which are created by several PFs. As a result of an old structural theorem of QFT, a PF is point-like local, if and only if it is a free field [53], i.e. if and only if the Fourier components $Z^\#(\theta)$ fulfil the free-field commutation relation which coalesce with those of the above ansatz for $S_{Z,Z} = 1 = S_{Z,Z^*}$. Although interacting PFs are necessarily nonlocal, it is an interesting question how nonlocal they must be in order not to fall under the reign of the structural theorem. It turns out that they can be localized in wedges but any sharper localization requirement reduces them to free fields. In the more *special context of the above ansatz* we find [52] the following proposition.

Proposition 2. *The requirement of wedge localization of a PF operator* $F(f) = \int F(x)f(x)\,\mathrm{d}^2x$, supp $f \in W$ *is equivalent to the Zamolodchikov–Faddeev structure of the* Z*-algebra. The corresponding* F *cannot be localized in smaller regions, i.e. the localization of* $F(f)$ *with* supp $f \in \mathcal{O} \subset W$ *is not in* $\mathcal{O}$ *but still uses all of* W.

Before doing the necessary calculation, let us put on record two more definitions of a general kind which are suggested by the proposition.

Definition 1. *We call PFs which generate the wedge algebra‡*

$$\mathcal{A}(W) = \mathrm{alg}\{F(\hat{f}), \forall f \text{ supp } \hat{f} \in W\}$$

PFG or one-particle polarization free wedge generators [52].

We omitted the w for wedge in our shorthand notation because wedges are the 'smallest' regions in Minkowski space which do not have the full space as the causal closure and possess PF. In view of the fact that we work more frequently in momentum space and its rapidity-parametrized mass-shell restriction (often referred to as one-particle wavefunctions), we reserve the simpler notation f without hat to the Fourier transforms.

Definition 2. *We call the improvement of localization obtained by intersecting* $\mathcal{A}(W)$ *for different wedges an improvement of 'quantum localization' [52], whereas the standard localization in* supp f *with the use of smeared out pointlike local fields* $A(f)$ *is referred to as classical (albeit in a* quantum *field theory).*

† Operators in QFT never possess analytic properties in x- or p-space. The notation and terminology in conformal field theory is a bit confusing on this point, because although it is used for operators it really should refer to vector states and expectation values in certain representations of the abstract operators. The use of modular methods require more conceptual clarity than standard methods.

‡ In this paper we do not discuss the necessity to distinguish between localized von Neumann algebras $\mathcal{A}(\mathcal{O})$ of bounded operators and polynomial algebras $\mathcal{P}(\mathcal{O})$ of affiliated unbounded operators as those formed from products of $F(f)$ and their precise relation.

We now prove the proposition by employing the so-called KMS condition for localized algebras. This property originally arose in thermal systems in cases where the thermodynamical limit for the infinitely extended system cannot be described in terms of a Gibbs formula (volume divergencies), but it later turned out to be generally valid for all systems which result von Neumann algebras $\mathcal{A}$ in a cyclic and separating state vector Ω:

$$(\Omega, A\sigma_t(B)\Omega) = (\Omega, \sigma_{t+i}(B)A\Omega) \tag{71}$$

where $\sigma_t(B) \equiv Ad\Delta^{it}(B)$ is the action of the modular group. Local algebras in QFT are known to have this commutation property with respect to the vacuum state at least as long as the localization region has a nontrivial causal complement, but they generally do not admit a natural thermodynamic limit description in terms of a sequence of increasing quantization boxes. For the wedge regions at hand, the localized field algebras are known to have the wedge affiliated Lorentz boost as their KMS automorphism group σ_t.

Proof. Consider first the KMS property of the two-point function

$$\langle F(f_1)F(f_2)\rangle = \langle F(f_2^{2\pi \mathrm{i}})F(f_1)\rangle = \langle F(f_2^{\pi \mathrm{i}})F(f_1^{-\mathrm{i}\pi})\rangle. \tag{72}$$

Rewritten in terms of f we have

$$\int f_1(\theta)\bar{f}_2(\theta)\,\mathrm{d}\theta = \int f_2(\theta - \mathrm{i}\pi)\bar{f}_1(\theta + \mathrm{i}\pi)\,\mathrm{d}\theta \tag{73}$$

which is an identity in view of the fact that the wedge supports properties for the test functions f together with their reality condition implying $f(\theta - \mathrm{i}\pi) = \bar{f}(\theta)$.

The four-point function $\langle 1, 2, 3, 4\rangle$ consists of three contributions, one from an intermediate vacuum state vector associated with the contraction scheme $\langle 12\rangle\langle 34\rangle$, another one from the direct intermediate two-particle contribution $\langle 14\rangle\langle 23\rangle$ and the third one from its exchanged (crossed) version $\langle 13\rangle\langle 24\rangle$. The latter is the only one which carries the interaction in form of the S-coefficients. In the would-be KMS relation

$$\begin{aligned} &\langle F(f_1)F(f_2)F(f_3)F(f_4)\rangle = \langle F(f_4^{-2\pi i})F(f_1)F(f_2)F(f_3)\rangle \\ &f^z(\theta) := \mathrm{a.c.} f|_{\theta\to\theta+z} \end{aligned} \tag{74}$$

the vacuum terms and the direct terms interchange their role on both sides of the equation and cancel out, whereas the crossed terms are related by analytic continuation. The required equality for the crossed term brings in the S-matrix via the relations (70) and yields

$$\begin{aligned} &\int\!\!\int \mathrm{d}\theta\,\mathrm{d}\theta'\, S(\theta-\theta')f_2(\theta)\bar{f}_4(\theta)f_1(\theta')\bar{f}_3(\theta') \\ &\qquad = \int\!\!\int \mathrm{d}\theta\,\mathrm{d}\theta'\, S(\theta-\theta')f_1(\theta)\bar{f}_3(\theta)f_4(\theta'-2\pi\mathrm{i})\bar{f}_2(\theta'). \end{aligned} \tag{75}$$

Again using the above boundary relation for the wavefunctions we rewrite the last product in the second line as $\bar{f}_4(\theta' - \mathrm{i}\pi)f_2(\theta' - \mathrm{i}\pi)$ and performing a contour shift $\theta' \to \theta' + \mathrm{i}\pi$, renaming $\theta \leftrightarrow \theta'$ and finally using the denseness of the wavefunctions in the Hilbert space, we obtain the crossing relation for S:

$$S(\theta) = S(-\theta + \mathrm{i}\pi). \tag{76}$$

Note that we already omitted the subscripts on S, since the identity $S_{Z,Z^*} = S_{Z,Z} \equiv S$ follows from the two different ways of calculating the crossed term, once by interchanging the two creation operators in $Z^*(\theta_3)Z^*(\theta_4)$ and then performing the direct contraction and another way by interchanging $Z(\theta_2)Z^*(\theta_3)$ and then being left with the vacuum contraction. Let us look at one more KMS relation for the six-point functions of the would-be PFG.

$$\langle F(f_1)\ldots F(f_6)\rangle = \langle F(f_6^{2\pi i})F(f_1)\ldots F(f_5)\rangle. \tag{77}$$

This time one has many more pairings. In fact, ordering with respect to pair contraction times four-point functions one may again group the various terms in those for which the pairing contraction is between adjacent Z and those where this only can be achieved by exchanges. The first group satisfies the KMS condition because of the previous verification for the two- and four-point functions. For the crossed contributions containing the wavefunctions say f_i and $\bar{f}_k$, those terms only agree on both sides after shifting upper C-contours into lower ones and vice versa. If S would contain poles in the physical sheet, then there are additional contributions and the KMS property only holds if these poles occur in symmetric pairs, i.e. in a crossing symmetric fashion. □

We will not pursue the fusion structure for the Z resulting from poles beyond noting that the particle spectrum already shows up in the fusion of the wedge-localized $Z(f)$. One of course expects agreement of the fusion structure of our PFGs with the formal Zamolodchikov conjecture†, however a detailed discussion of fusion would go beyond the aim of this paper and will be the subject of a separate paper. It should be stressed that the simple quantum mechanical picture of fusion in terms of bound states only holds for the above model with pair interactions and not for more realistic models with real (on-shell) particle creation. All models whether they are real particle conserving or not (except free fields) have a rich virtual particle structure (as will be shown later), i.e. the particle content of operators A with compact localization, e.g. $A \in \mathcal{A}(\mathcal{O})$ complies with the 'folklore' that all particle matrix elements

$$^{out}\langle p_1, \ldots, p_k|A|q_1, \ldots, q_l\rangle^{in} \neq 0 \tag{78}$$

as long as they are not forced to vanish by superselection rules.

Although we have explained the basic concepts in the case of diagonal S-coefficients in the Z-algebra, one realizes immediately that one can generalize the formalism to *matrix-valued* 'pair interactions' S. The operator formalism (the associativity) then leads to the Yang–Baxter conditions and the crossing relations are again equivalent to the KMS property for the wedge generators $F(f)$.

The relation of the above observation with local quantum physics (LQP) becomes tighter, if one remembers that the Lorentz boost, which featured in the above KMS condition, also appears together with the TCP operator in the Tomita modular theory for the pair $(\mathcal{A}(W), \Omega)$:

$$S_T A\Omega = A^*\Omega \qquad A \in \mathcal{A}(W) \tag{79}$$

which defines the antilinear, unbounded, closable, involutive (on its domain) Tomita operator S_T. Its polar decomposition

$$S_T = J\Delta^{\frac{1}{2}} \tag{80}$$

defines a positive unbounded $\Delta^{\frac{1}{2}}$ and an antiunitary involutive J and the nontrivial part of Tomita's theorem (with improvements by Takesaki) is that the unitity Δ^{it} defines an automorphism of the algebra i.e. $\sigma_t(\mathcal{A}) \equiv \Delta^{it}\mathcal{A}\Delta^{-it} = \mathcal{A}$ and the J maps into antiunitarily into its commutant $j(\mathcal{A}) \equiv J\mathcal{A}J = \mathcal{A}'$. The wedge situation is a special illustration for the Tomita theory. In that case both operators are well known; the modular group is the one-parametric wedge affiliated Lorentz boost group $\Delta^{it} = U(\Lambda(-2\pi t))$, and the J in $d = 1 + 1$ LQPs is the fundamental TCP-operator (in higher dimensions it is only different by a π-rotation around the spatial wedge axis). The prerequisite for the general Tomita situation is that the vector in the pair (algebra, vector) is cyclic and separating (no annihilation operators in the von

† In fact it is only through the PFG's $F(x)$ that the Z–F algebra and the fusion rules for the Z receive a space-time interpretation. The close relation to a kind of relativistic QM only happens on the level of wedge localization; the algebras resulting from intersections of wedge algebras loose this quantum mechanical aspect and show the full virtual particle creation/annihilation polarization structure.

Neumann algebra resp. cyclicity of its commutant relative to the reference vector). In LQP these properties are guaranteed for localization regions $\mathcal{O}$ with nontrivial causal complement $\mathcal{O}'$ thanks to the Reeh–Schlieder theorem. Returning to our wedge situation we conclude from the Bisognano–Wichmann result that the commutant of $\mathcal{A}(W)$ is geometric, i.e. it fulfils Haag duality $\mathcal{A}(W)' = \mathcal{A}(W')$, a fact which can be shown to be modified by Klein factors in J in case of deviation from Bose statistics.

There is one more structural element following from 'quantum localization' beyond wedge localization.

Proposition 3. *Operators localized in double cones $A \in A(\mathcal{O})$ obey a recursion relation in their expansion coefficients in terms of PFG operators*

$$A = \sum \frac{1}{n!} \int_C \ldots \int_C a_n(\theta_1, \ldots \theta_n) : Z(\theta_1) \ldots Z(\theta_n) : \mathrm{d}\theta_1 \ldots \mathrm{d}\theta_n$$

$$= \sum \frac{1}{n!} \int \ldots \int \hat{a}_n(x_1, \ldots x_n) : F(x_1) \ldots F(x_n) : \mathrm{d}^2x_1 \ldots \mathrm{d}^2x_n$$

$$\mathrm{supp}\, \hat{a} \in W^{\otimes n}$$

$$\mathrm{i} \lim_{\theta \to \theta_1} (\theta - \theta_1) a_{n+1}(\theta, \theta_1, \ldots, \theta_n) = \left(1 - \prod_{i=2}^{n} S(\theta_1 - \theta_i)\right) a_{n-1}(\theta_2, \ldots, \theta_n).$$

Remark 1. *In order to compare (see below) with Smirnov's [54] axioms we wrote the recursion in rapidity space instead of in x-space light-ray restriction which would be more physical and natural to our modular approach. The series extends typically to infinity. Only for special operators (e.g. bilinears as the energy momentum tensor) in special models with rapidity independent S-matrices (e.g. Ising, Federbush) for which the bracket involving the product of two-particle S-matrices vanishes, the series restricts to a polynomial expression in Z. Therefore apart from these special cases, an operator $A \in \mathcal{A}(\mathcal{O})$ with $a_1 \neq 0$ applied to the vacuum creates a one-particle component which an admixture of an infinite cloud of additional particles (particle–antiparticle polarization cloud). The above recursion together with Payley–Wiener-type bounds for the increase of the a_n in imaginary θ-directions (depending on the shape and size of $\mathcal{O}$).*

The proof follows rather straightforwardly from the quantum localization idea

$$\mathcal{A}(\mathcal{O}) = [U(a)\mathcal{A}(W)U^{-1}(a)]' \cap \mathcal{A}(W) \tag{81}$$

i.e. we are considering the relative commutant inside the wedge algebra. Using the PFG's $F(f)$, the $A \in \mathcal{A}(\mathcal{O})$ are characterized by [52]

$$[A, F(\hat{f}_a)] = 0 \qquad \forall \hat{f} \in W \tag{82}$$

where $\hat{f}_a(x) = \hat{f}(x - a)$, $a \in W$. One immediately realizes that the contribution of the commutator to the nth power in F yields a relation between the a_{n-1} and a_{n+1} (from the creation/annihilation part of $F(\hat{f}_a)$). The details of this relation are easier, if one passes to the light-ray restriction which in the present approach turns out to be a very nontrivial result of modular theory [52, 55, 56].

Proposition 4. *The relative commutant for light-like translations with $a_+ = (1, 1)$ defines a 'satellite' chiral conformal field theory via the (half) net on the (upper) + light ray*

$$\mathcal{A}(I_{a,\mathrm{e}^{2\pi t}+a}) = U(a, a)\Delta^{-it} \left(\mathcal{A}(W_{a_+})' \cap \mathcal{A}(W)\right) \Delta^{it} U^{-1}(a, a) \tag{83}$$

where $I_{a,b}$ with $b > a \geqslant 0$ denotes an interval on the right upper light ray. This net is cyclic and separating with respect to the vacuum in the reduced Hilbert space

$$\begin{aligned} &H_+ = \overline{\mathcal{M}_+\Omega} = P_+H \subset H = \overline{\mathcal{A}(W)\Omega} \\ &\mathcal{M}_+ \equiv \cup_t A(I_{0,\mathrm{e}^{2\pi t}}) \qquad E_+(\mathcal{A}(W)) = \mathcal{M}_+ = P_+\mathcal{A}(W)P_+ \end{aligned} \tag{84}$$

where the last relation defines a conditional expectation. The application of J to gives the left lower part of this light ray which is needed for the full net.

Remark 2. *The most surprising aspect of this proposition is that this light-ray affiliated chiral conformal theory exhibits the 'blow-up' property, i.e. can be activated to reconstitute the two-dimensional net by association of the $-$ light-ray translation*

$$\begin{aligned}&\mathcal{A}(W) = \mathrm{alg} \cup_{a>0} \{\mathcal{M}_+, U_-(a)\} \\ &\mathcal{A} = \mathcal{A}(W) \vee \mathcal{A}(W)'.\end{aligned} \tag{85}$$

The Moebius groups $SL(2,R)_\pm$ account for six parameters in contradistinction to the three parameters of the two-dimensional Poincaré group of the massive theory. Most of the former are 'hidden' and the original theory perceives these additional symmetries only in its $P_\pm$ projections (for the proofs see [52,56]).

The light-ray reduction reduces the derivation of the recursion relation to a one-dimensional LQP problem and the reader may carry out the missing algebra without much effort. This reduction also helps significantly in the demonstration that the $\mathcal{A}(\mathcal{O})$ spaces are nontrivial, i.e. contain more elements than multiples of the identity. It is a fascinating experience to see that the existence problem for nontrivial QFTs (which in the quantization approach always pointed in the direction of getting good short-distance properties and in particular the renormalizability requirement $\dim \mathcal{L}_{int} \leqslant$ dim space-time), in the modular approach, which does not use individual 'field-coordinatizations', relates the existence of nontrivial field theories associated with interacting PFGs to the nontriviality of intersections which represent double-cone algebras. The above constructions only determine operators in the sense of bilinear forms.

At this point it is appropriate to address the question of what we learned from this approach as compared with the Karowski–Weisz–Smirnov 'axiomatics' [53]†. Actually, a considerable part of that axiomatics has been reduced to specializations of general field-theoretic properties within the LSZ framework [50], apart from the algebraic and analytic aspects of the fundamental crossing property. Since the LSZ formalism itself can be derived from the basic causality and spectral properties of say Wightman QFT, one may even want to have a more direct physical understanding of the other properties. This is achieved by realizing that the a_n-coefficients have the interpretation of the connected part of formfactors of A, for selfconjugate models

$$\begin{aligned}&a_n(\theta_1,\ldots,\theta_n) = \langle\Omega|A|\theta_1,\ldots,\theta_n\rangle^{in} \\ &\theta_1 < \theta_2 < \cdots < \theta_n\end{aligned} \tag{86}$$

$$a_n(\theta_1,\ldots,\theta_\nu,\theta_{\nu+1}-\mathrm{i}\pi,\ldots,\theta_n-\mathrm{i}\pi) = {}^{out}\langle\theta_1,\ldots,\theta_\nu|A|\theta_{\nu+1},\ldots;\theta_n\rangle^{in}_{conn}. \tag{87}$$

The relations for different orderings of θ follows from the algebraic structures of the Z.

In the diagonal case this connection between Z and in- and out-creation/annihilation operators can be seen directly via representing the Z in a bosonic/fermionic Fock space of the incoming particles in the form

$$Z(\theta) = a_{in}(\theta)\mathrm{e}^{\mathrm{i}\int a^*_{in}(\theta)a(\theta)\,\mathrm{d}\theta}. \tag{88}$$

However, such representations are not known for the nondiagonal case. But once one obtained the double-cone localized operators the theory itself (scattering theory as a consequence of the locality + spectral structure) assures the existence of Z in terms of incoming particle creation/annihilation operators, albeit not in terms of simple exponential formulae.

The modular theory for wedges in terms of PFGs really explains the KWS axiomatics by integrating it back into the fundamental principles of general QFT. In particular, the notoriously

† Our operator notation is closer to Lashkevich 1994 [54].

difficult crossing symmetry for the first time finds its deeper explanation in Hawking–Unruh thermal KMS properties once one realizes that a curved space-time Killing vector (a classical concept) is not as important quantum localization of operator algebras. With these remarks we have achieved our goal of deriving and explaining all axioms of the KWS approach in terms of localization properties of PFGs with pair interactions.

This raises the question if the PFG's $F(x)$ in their property as wedge algebra generators, could not also exist for higher dimensions. In that case their application more than one time to the vacuum would generate a state whose particle content (the real particle structure) is already very complicated. As is often the case in general QFT, it is easier to see what does not work, i.e. to prove no-go theorems. Indeed, if the interacting PFGs exist at all, their causally closed living space $\mathcal{O}$ cannot be (even a tiny little bit) smaller than a wedge $\mathcal{O} \subset W$. As was already stated at the beginning, if there would be space-like directions with an arbitrarily small conic surrounding which are contained in W but not in $\mathcal{O}$, it is fairly easy to generalize the proof† of the Jost–Schroer theorem and show that the commutators of such PFGs must be a c-number which is determined by their two-point function. However, the method used in those No-Go theorems has no extension to the wedge region. If wedge algebras can indeed be generated by PFGs, one expects again that modular theory does not only relate them to the S-matrix so that their correlations can be expressed in terms of products of S-matrix elements (with partial summations reminiscent of inclusive processes) and furthermore that the mysterious crossing symmetries for the S-matrix and formfactors find their explanation in the thermal KMS properties. This surprising relation between particle physics and the thermal properties of Hawking–Unruh wedge horizons has attracted the attention of many physicists, the ideas most close to those of the present work and several older papers [46] of the present author are those in [49]. However, it should be clear that as long as higher-dimensional PFGs have yet to be constructed or at least their existence established, the mediators between off- and on-shell are still missing and there is no proof beyond the one for factorizing models presented before.

There is also an interesting extension of the KWS axiomatics in the form of a pair of satellite chiral conformal theories. In contradistinction to the standard short-distance association the light-ray association via modular theory is not just a one-way street; the blow-up property with the help of adjoining the opposite light-cone translation allows us to return, so that hidden conformal symmetries become relevant for the massive theory or more precisely for the massive theory projected into the $H_\pm$ subspaces.

Note that the present construction principle can be directly used for the systematic construction of chiral conformal theories. For the construction of W-like algebras one starts with PFG generators on a half-line. Modular theory assures us that in principle every system of S-coefficients fulfilling the Z–F algebra leads to a bosonic/fermionic conformal theory granted that the previous relative commutator algebra is nontrivial. This is a construction scheme which could not have been guessed within the framework of point-like fields.

Another apparently simple but untested idea suggested by the present concepts is the classification of wedge algebras with nongeometric commutator algebras via statistics Klein factors or constant S-matrices in J. Examples are the Ising field theory and the order/disorder fields. For the more interesting case of plektonic R-matrices which appear in the exchange algebras [58] of charge carrying fields, one knows that these algebras in contradistinction to bosonic/fermionic (e.g. W-algebras) are incomplete since the distributional character at coalescent points is left unspecified. This is not the case if one uses the R-data as an input into plektonic $Z^{\#}(\theta)$. The Hilbert space obtained by iterative application of Z-creation operators is not compatible with a Fock space structure. Rather, the n-particle subspace has the structure of a

† I learned this from D Buchholz.

path space as known from the representation theory of intertwiner algebras. The combinatorial complications should be offset by the simplicity of constant S-matrices. As the operator representation of the massive Ising model shows, the constant S case should even have a simple coefficient series in the massive case.

6. Concluding remarks

Whereas causality and locality principles used to play an important role in the past (the LSZ framework, the Kramers–Kronig relations in high-energy physics and their experimental check in high-energy nucleon scattering), they have been less prominent in the more global functional integral formulation of QFT. In S-matrix models as the Veneziano dual model the role of these principles is even harder to see, but the idea that crossing symmetry which underlies duality is a deep on-shell manifestation of causality always carried a lot of plausibility. The difficulty here is that crossing symmetry was primarily an observation on Feynman diagrams whose relation to the causality- and particle-structure was never clarified as that of other symmetries, e.g. as happened with the simpler TCP symmetry. In fact, the dual model which was originally intended to probe the structure of a nonperturbative S-matrix and to shed light on the elusive crossing symmetry, was soon treated as a separate issue with the original QFT motivation being forgotten. After several abrupt changes of interpretation and also finally of the mathematical formalism (the so-called 'string revolutions') it finally reached its present form of string theory with interesting mathematical connections but without convincing conceptual content. The status of locality within interacting string theory is unknown (the answer one gets depends on the person one asks†). If the word string could be interpreted as indicating a space-time localization and not just referring to certain spectral properties, then it would be part of local QFT and all the structural statements in this article would immediately be applicable. However, in this case it should be possible to have an intrinsic formulation (say, analogous to the Wightman framework). As it stands now, string theory is synonymous with a collection of computational steps. Related to this is the total lack of an answer to the question: what physical principle is it which asks for a string-like extension in order to be realized? One should like to have a physically more compelling reason than just saying that after having been interested for many years in point-like fields one wants to study string-like extensions.

The development of physical theories has been (and still is in my opinion) the unfolding of ever more general realizations of physical principles. For example, the semi-infinite string-like localization of $d = 2 + 1$ anyons/plektons or topological charges (in the sense of algebraic QFT [1]) is required by the more general realization of causality; if one allows only compact extensions, one would fall back on bosons/fermions and ordinary charges. Most structural properties in LQP have been understood as an unfolding of realizations of physical principles. One hopes that this fruitful viewpoint of this century may not get completely lost in the ongoing process of marketing and globalization in the production of publications which is taking place at the end of it.

Note added in proof. There have been very interesting new results by modular methods [59].

Acknowledgment

I am indebted to H-W Wiesbrock for discussions and for critical reading of the manuscript.

† Part of the problem may originate from the fact that quantum causality and locality is often confused with support or geometrical properties of Lagrangians, one of the negative side effects of the naive interpretation of Euclidean field theory.

References

[1] Haag R 1992 *Local Quantum Physics* (Berlin: Springer)
[2] Lee T D and Wick G C 1969 *Nucl. Phys.* B **9** 209
[3] Marques G C and Swieca J A 1972 *Nucl. Phys.* B **43** 205
[4] Doplicher S, Fredenhagen K and Roberts J E 1995 *Commun. Math. Phys.* **152** 187
[5] Amati D, Ciafaloni M and Veneziano G 1990 *Nucl. Phys.* B **347** 551
See also Kempf A 1997 *Proc. 5th Wigner Symposium (Vienna, August 1997)*
[6] Schroer B 1998 Localization and nonperturbative local quantum physics *Preprint* hep-th/9805093
[7] Roberts J H 1976 Spontaneously broken gauge symmetries and superselection rules *Proc. Int. School of Mathematical Physics* ed G Galavotti (University de Camerino)
[8] Newton T D and Wigner E P 1949 *Rev. Mod. Phys.* **21** 400
[9] Sewell G L 1982 *Ann. Phys., Lpz.* **141** 201
Summers S and Verch R 1996 *Lett. Math. Phys.* **37** 145
[10] Schroer B 1999 *Ann. Phys., Lpz.* **275** to appear
(Schroer B 1997 *Preprint* hep-th/9712124)
[11] Jaekel C 1998 Decay of spatial correlations in thermal states *Preprint* hep-th/9803233
[12] Weinberg S 1955 *The Quantum Theory of Fields I: Foundations* (Cambridge: Cambridge University Press)
[13] Buchholz D 1994 On the manifestation of particles *Mathematical Physics Towards the 21st Century* ed R Sen and A Gersten (Beer-Sheva: Ben Gurion University Press) p 177
[14] Bratteli O and Robinson D W 1979 *Operator Algebras and Quantum Statistical Mechanics* vols 1 and 2 (Berlin: Springer) and references therein
[15] Haag R and Schroer B 1962 *J. Math. Phys.* **3** 248
[16] Baumgärtel H, Jurke M and Lledó F 1995 *Rep. Math. Phys.* **35** 101–27
[17] Bisognano J J and Wichmann E H 1975 *J. Math. Phys.* **16** 985
Bisognano J J and Wichmann E H 1976 *J. Math. Phys.* **17** 303
[18] Buchholz D and Yngvason J 1994 *Phys. Rev. Lett.* **73** 613
[19] Buchholz D and Wichmann E H 1986 *Commun. Math. Phys.* **106** 321
[20] Bogoliubov N N and Shirkov D V 1993 *Introduction to the Theory of Quantized Fields* (New York) and references therein
[21] Epstein H and Glaser V 1973 *Ann. Inst. Henri Poincaré* A **19** 211
[22] Duetsch M and Fredenhagen K 1998 A local (perturbative) construction of observables in gauge theories *Preprint* hep-th/9807078
[23] Brunetti R and Fredenhagen K 1997 Interacting field theories in curved space: renormalization of ϕ^4 *Preprint* gr-qc/970148
[24] Becchi C, Rouet A and Stora R 1975 *Commun. Math. Phys.* **42** 127
Becchi C, Rouet A and Stora R 1976 *Ann. Phys., NY* **98** 287
[25] Kugo T and Ojima I 1979 *Suppl. Prog. Theor. Phys.* **66** 1
[26] Duetsch M, Hurth T, Krahe F and Scharf G 1993 *Nuovo Cimento* A **160** 1029
Aste A and Scharf G 1998 *Preprint* hep-th/9803011
Grigore D 1998 On the uniqueness of the non-Abelian gauge theories in epstein-glaser approach to gauge theories *Preprint* hep-th/9806244
[27] Dubois-Violette M unpublished
[28] Duetsch M and Schroer B 1999 Massive vector mesons and gauge theories *Preprint* hep-th/9906089
[29] Schwinger J 1962 IAE Report (Vienna) *Lectures at the Seminar on Theoretical Physics (Trieste, 1962)*
Schwinger J 1962 *Phys. Rev.* **128** 2425
Swieca J A 1976 *Phys. Rev.* D **13** 312
[30] Virasoro M A 1970 *Phys. Rev.* D **1** 2933
[31] Longo R and Rehren K-H 1995 *Rev. Math. Phys.* **7** 567–98
[32] Schroer B and Wiesbrock H-W 1999 *Rev. Mod. Phys.* to appear
(Schroer B and Wiesbrock H-W 1998 Modular theory and geometry *Preprint* math-ph/9809003)
[33] Lowenstein J H and Schroer B 1971 *Phys. Rev.* D **3** 1981
[34] Lowenstein J H 1967 *Commun. Math. Phys.* **6** 49 and references therein
[35] Ramond P 1971 *Phys. Rev.* D **3** 2415
[36] Schroer B 1974 A trip to scalingland *V Brazilian Symp. Theoretical Physics (Rio de Janeiro, January 1974)* vol 1, ed E Ferreira (Livros Técnicos e Cientificos Editora S.A.) a scanned version can be found in [6]
[37] Schroer B and Swieca J A 1974 *Phys. Rev.* D **10** 480
Schroer B, Swieca J A and Völkel A H 1975 *Phys. Rev.* D **11** 1509

[38] Belavin A A, Polyakov A M and Zamolodchikov A B 1984 *Nucl. Phys.* B **247** 83
[39] Borchers H J 1992 *Commun. Math. Phys.* **143** 315
Borchers H J 1995 *Ann. Inst. Henri Poincaré* **63** 331
[40] Wiesbrock H-W 1997 *Commun. Math. Phys.* **157** 83
Wiesbrock H-W 1997 *Commun. Math. Phys.* **184** 683 (erratum)
Wiesbrock H-W 1997 *Lett. Math. Phys.* **39** 203
[41] Araki H and Zsido L 1995 *An Extension of the Structure Theorem of Borchers with an Application to Half-Sided Modular Inclusions* preliminary version
[42] Borchers H-J and Yngvason J 1998 Modular groups of quantum fields in thermal states *Preprint*
[43] Wassermann A 1997 Operator algebras and conformal field theory *University of Cambridge Preprint*
[44] Takesaki M 1972 Conditional expectations in von Neumann algebras *J. Funct. Anal.* **9** 306
[45] Kawahigashi Y, Longo R and Mueger M 1999 Multi-interval subfactors and modularity of representations in conformal field theories *Preprint* math.OA/9903104
[46] Schroer B 1997 *Ann. Phys., Lpz.* **255** 270
[47] Fredenhagen K and Joerss 1996 *Commun. Math. Phys.* **176** 541
[48] Buchholz D, Dreyer O, Florig M and Summers S 1998 Geometric modular action and space-time symmetry groups *Preprint* math.ph./9805026
[49] Niedermaier M 1998 *Commun. Math. Phys.* **196** 411
[50] Babujian H, Fring A, Karowski M and Zapletal A 1998 Exact form factors in integrable quantum field theories: the sine–Gordon model *Nucl. Phys.* B **538** 535
[51] Wightman A S 1966 High energy interactions and field theory *Cargese Lectures in Theoretical Physics 1964* (New York: Gordon and Breach)
[52] Schroer B and Wiesbrock H-W 1999 *Rev. Math. Phys.* to appear
(Schroer B and Wiesbrock H-W 1998 Modular constructions of quantum field theories with interactions *Preprint* hep-th/9812251)
[53] Karowski M and Weisz P 1978 *Nucl. Phys.* B **139** 445
[54] Smirnov F A 1992 Form factors in completely integrable models *Adv. Series in Math. Phys. 14* (Singapore: World Scientific)
Lashkevich M Yu 1994 Sectors of mutually local fields in integrable models of quantum field theory *Preprint* hep-th/9406118
[55] Schroer B and Wiesbrock H-W 1999 *Rev. Math. Phys.* to appear
(Schroer B and Wiesbrock H-W 1999 Looking beyond the thermal horizon: hidden symmetries in chiral models *Preprint* hep-th/9901031)
[56] Guido D, Longo R, Roberts J E and Verch R 1999 Charged sectors, spin and statistics in quantum field theory on curved spacetimes *Preprint* math-ph/9906019
[57] Guido D, Longo R and Wiesbrock H-W 1998 *Commun. Math. Phys.* **192** 217
[58] Rehren K H and Schroer B 1989 *Nucl. Phys.* B **312** 715
[59] Schroer B 1999 Modular theory and Eyvind Wichmann's contributions to modern particle physics theory *Preprint* hep-th/9906071