FRONTIERS IN PHYSICS - RISING STARS

EDITED BY: Alex Hansen, Ewald Moser, Matjaž Perc, Lorenzo Pavesi, Rudolf von Steiger, Nicholas X. Fang, J. W. F. Valle, Jan De Boer, Christian F. Klingenberg, Laura Elisa Marcucci, Jasper Van Der Gucht and Alexandre M. Zagoskin

PUBLISHED IN: Frontiers in Physics

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88971-105-5
DOI 10.3389/978-2-88971-105-5

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

FRONTIERS IN PHYSICS - RISING STARS

Citation: Hansen, A., Moser, E., Perc, M., Pavesi, L., von Steiger, R., Fang, N. X., Valle, J. W. F., De Boer, J., Klingenberg, C. F., Marcucci, L. E., Van Der Gucht, J., Zagoskin, A. M., eds. (2021). Frontiers in Physics - Rising Stars. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-105-5

Table of Contents

SOFT MATTER PHYSICS

SPACE PHYSICS

COMPUTATIONAL PHYSICS

Andrew Ross Winters obtained his PhD in 2014 from Prof. Kopriva at Florida State University, USA on the topic of discontinuous Galerkin methods. Afterwards he joined Gregor Gassner's research group in Cologne, Germany for 5 years. He then obtained a permanent position at Linköping University in Sweden, where he is working at present.

Codes with a high level of parallelization are imminently suited to the present generations of super computers. In computational fluid mechanics the discontinuous Galerkin numerical methods fit extremely well into this paradigm. This very timely article gives an excellent overview of this method, taking into account important recent developments, that make this method even more useful.

frontiers in Physics

REVIEW
published: 29 January 2021
doi: 10.3389/fphy.2020.500690

A Novel Robust Strategy for Discontinuous Galerkin Methods in Computational Fluid Mechanics: Why? When? What? Where?

Gregor J. Gassner[1] *and Andrew R. Winters*[2]*

[1]Division of Mathematics, Department of Mathematics and Computer Science, Center for Data and Simulation Science, University of Cologne, Cologne, Germany, [2]Division of Computational Mathematics, Department of Mathematics, Linköping University, Linköping, Sweden

In this paper we will review a recent emerging paradigm shift in the construction and analysis of high order Discontinuous Galerkin methods applied to approximate solutions of hyperbolic or mixed hyperbolic-parabolic partial differential equations (PDEs) in computational physics. There is a long history using DG methods to approximate the solution of partial differential equations in computational physics with successful applications in linear wave propagation, like those governed by Maxwell's equations, incompressible and compressible fluid and plasma dynamics governed by the Navier-Stokes and the Magnetohydrodynamics equations, or as a solver for ordinary differential equations (ODEs), e.g., in structural mechanics. The DG method amalgamates ideas from several existing methods such as the Finite Element Galerkin method (FEM) and the Finite Volume method (FVM) and is specifically applied to problems with advection dominated properties, such as fast moving fluids or wave propagation. In the numerics community, DG methods are infamous for being computationally complex and, due to their high order nature, as having issues with robustness, i.e., these methods are sometimes prone to crashing easily. In this article we will focus on efficient nodal versions of the DG scheme and present recent ideas to restore its robustness, its connections to and influence by other sectors of the numerical community, such as the finite difference community, and further discuss this young, but rapidly developing research topic by highlighting the main contributions and a closing discussion about possible next lines of research.

Keywords: discontinuous Galerkin method, robustness, split form, dealiasing, summation-by-parts, second law of thermodynamics, entropy stability

OPEN ACCESS

Edited by:
Christian F. Klingenberg, Julius Maximilian University of Würzburg, Germany

Reviewed by:
Hendrik Ranocha, King Abdullah University of Science and Technology, Saudi Arabia
Francesco Fambri, Max Planck Institute for Plasma Physics (IPP), Germany

***Correspondence:**
Andrew R. Winters
andrew.ross.winters@liu.se

Specialty section:
This article was submitted to Computational Physics, a section of the journal Frontiers in Physics

Received: *01 October 2019*
Accepted: *26 November 2020*
Published: *29 January 2021*

Citation:
Gassner GJ and Winters AR (2021) A Novel Robust Strategy for Discontinuous Galerkin Methods in Computational Fluid Mechanics: Why? When? What? Where?. Front. Phys. 8:500690. doi: 10.3389/fphy.2020.500690

1 A BRIEF INTRODUCTION TO DG

The first discontinuous Galerkin (DG) type discretisation is either attributed to Reed and Hill in 1973 [1] for an application to steady state scalar hyperbolic linear advection to model neutron transport, or to Nitsche in 1971 [2] who introduced a discontinuous finite element method (FEM) to solve elliptic problems with non-conforming approximation spaces. It was however a series of papers by Cockburn and Shu et al. starting 20 years later [3–6] that introduced the *modern* form of the so-called Runge-Kutta DG scheme. They applied the method especially to nonlinear hyperbolic problems such as the compressible Euler equations on unstructured simplex grids with slope limiting to capture shocks. Bassi and Rebay were the first that extended the DG method to the compressible

Navier-Stokes equations [7]. They used a fully discontinuous ansatz based on a mixed variational formulation, where they rewrote the second order partial differential equation (PDE) into a first order system. The resulting DG formulation requires numerical fluxes for the advective as well as for the diffusive part. Although the methods gave reasonable results for the compressible Navier-Stokes equations, an analysis of the method in Arnold and Brezzi et al. [8, 9] applied to pure elliptic problems revealed how to improve the method in terms of convergence rate, adjoint consistency, and stability. Since the introduction of its modern form, the DG method has been applied and advanced by many researchers across different scientific disciplines around the world. The DG method is used in a wide range of applications such as compressible flows [10–12], electromagnetics and optics [13–16], acoustics [17–21], meteorology [22–25], and geophysics [26, 27]. The first book available on DG was basically a collection of papers [28]. Since then, many different text books on DG are available focusing on theoretical developments as well as specific implementation details, e.g., [29–31].

One of the main applications of DG methods is the discretisation of nonlinear advection-diffusion problems of the form

$$\mathbf{u}_t + \overrightarrow{\nabla}_x \cdot \overleftrightarrow{\mathbf{f}}(\mathbf{u}) = \overrightarrow{\nabla}_x \cdot \overleftrightarrow{\mathbf{f}}_v\left(\mathbf{u}, \overrightarrow{\nabla}_x \mathbf{u}\right), \tag{1}$$

where $\mathbf{u}$ is the vector of conserved quantities, e.g., the mass, momentum, or energy. The vector $\mathbf{f}(\mathbf{u})$ defines the flux functions that in general depend nonlinearly on the solution $\mathbf{u}$, and can be compactly written with the double arrow notation as block vectors

$$\overleftrightarrow{\mathbf{f}} = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_3 \end{bmatrix}, \tag{2}$$

with the fluxes $\mathbf{f}_i$ in each spatial direction x_i, $i = 1,2,3$. The viscous fluxes are denoted by $\mathbf{f}_v$ and not only depend on the solution, but also on its spatial gradient

$$\overrightarrow{\nabla}_x \mathbf{u} = \begin{bmatrix} \mathbf{u}_x \\ \mathbf{u}_y \\ \mathbf{u}_z \end{bmatrix}, \tag{3}$$

thus modeling parabolic effects, e.g., heat conduction. The problem is typically defined on a given spatial domain $\Omega \subset \mathbb{R}^3$, with a final time T, and suitable initial and boundary conditions.

The DG scheme is based on a Galerkin type weak formulation. For the sake of simplicity, we drop the viscous second order terms in what follows and refer to, e.g. [32], for a complete description of the advection-diffusion case. To construct the approximation space of the DG method, the domain is split into non-overlapping elements $E \subset \Omega$. Each component of the solution $\mathbf{u}$ is represented as a polynomial function inside each element

$$\mathbf{u}\left(\overrightarrow{x}, t\right)\Big|_E \approx \mathbf{U}\left(\overrightarrow{x}, t\right) = \sum_{j=0}^{P(N)} \mathbf{U}_j(t)\, \phi_j\left(\overrightarrow{x}\right), \tag{4}$$

where $P(N)$ is the number of polynomial basis functions depending on the polynomial degree N. The time dependent polynomial coefficients are $\mathbf{U}_j(t)$, and $\phi_j(\overrightarrow{x})$ spans the polynomial basis. The DG approximation space is polynomial inside an element, but discontinuous across element interfaces. For a given element E, we define first the inner product for state vectors

$$\langle \mathbf{u}, \mathbf{v} \rangle_E = \int_E \mathbf{u}^T \mathbf{v}\, \mathrm{d}\overrightarrow{x}. \tag{5}$$

Similarly, for block vectors,

$$\langle \overleftrightarrow{\mathbf{f}}, \overleftrightarrow{\mathbf{g}} \rangle_E = \int_E \sum_{i=1}^{3} \mathbf{f}_i^T \mathbf{g}_i\, \mathrm{d}\overrightarrow{x}. \tag{6}$$

We obtain the weak formulation by multiplying each equation by a polynomial test function $\phi(\overrightarrow{x})$. Next, we integrate over the element E and use integration-by-parts to move the spatial derivatives off of the physical fluxes onto the test function

$$\langle \mathbf{U}_t, \phi \rangle_E + \oint_{\partial E} \phi^T \overleftrightarrow{\mathbf{f}} \cdot \overrightarrow{n}\, \mathrm{dS} - \langle \overleftrightarrow{\mathbf{f}}, \overrightarrow{\nabla}_x \phi \rangle_E = 0. \tag{7}$$

If we choose the test function ϕ to be all the polynomial basis functions from the solution ansatz space $\{\phi_j\}_{j=0}^{P}$ it generates P equations in each element for each state variable. This matches exactly the number of unknown polynomial coefficients $\mathbf{U}_j(t)$. Due to the discontinuous ansatz, the flux normal $\mathbf{f} \cdot \overrightarrow{n}$ at the surface integral is not uniquely defined. Borrowing ideas from the Finite Volume (FV) community, this non-unique normal flux function is replaced and approximated with a so-called numerical surface flux function

$$\overleftrightarrow{\mathbf{f}} \cdot \overrightarrow{n} \approx \mathbf{f}^*(\mathbf{U}^+, \mathbf{U}), \tag{8}$$

that depends on the two values at the element interface, i.e. on the value inside the element $\mathbf{U}$ and outside from the neighbor element $\mathbf{U}^+$. Typically, the numerical surface fluxes are constructed from (approximate) Riemann solvers, e.g., [33]. With the surface numerical flux, we arrive at the semi-discrete DG formulation of the advection problem in weak form

$$\langle \mathbf{U}_t, \phi \rangle_E + \oint_{\partial E} \phi^T \mathbf{f}^*(\mathbf{U}^+, \mathbf{U})\, \mathrm{dS} - \langle \overleftrightarrow{\mathbf{f}}, \overrightarrow{\nabla}_x \phi \rangle_E = 0, \tag{9}$$

or if we apply integration-by-parts once more to the volume terms it becomes the so-called strong DG formulation

$$\langle \mathbf{U}_t, \phi \rangle_E + \oint_{\partial E} \phi^T \left(\mathbf{f}^*(\mathbf{U}^+, \mathbf{U}) - \mathbf{f} \cdot \overrightarrow{n} \right) \mathrm{dS} + \langle \overrightarrow{\nabla}_x \cdot \overleftrightarrow{\mathbf{f}}, \phi \rangle_E = 0, \tag{10}$$

where the surface contribution is a penalty between the numerical surface flux and the normal flux evaluated from the interior of an element. Note, the weak and strong form DG formulations are equivalent [34].

There are still many critical decisions necessary before the DG formulation produces an algorithm that can be implemented. The

type of element shape needs to be decided as well as which polynomial basis. For example, modal polynomial basis functions for tetrahedral elements or nodal tensor product polynomials for hexahedral elements. In addition, the surface integral and the volume integral needs to be discretized. In most cases, the integrals are approximated with numerical quadrature rules, e.g., high order Gauss-Legendre quadrature and cubature. Many variants and detailed descriptions can be found in text books on DG methods and their implementation, e.g., [28–31]. It is important to note that these choices involving the element type, basis functions, and approximation of inner products all have a major impact on the performance of the resulting DG scheme in terms of computational complexity and robustness due, e.g., to the presence of spurious oscillations near discontinuities that result in unphysical solution states (like negative density or pressure) or aliasing instabilities. Many mechanisms exist in the DG community to combat spurious oscillations (i.e., shock capturing) such as slope [3, 5, 35] or WENO [36, 37] limiters, filtering [29, 38, 39], finite volume sub-cells [40–42], MOOD-type limiting [43–47], or artificial viscosity [48, 49]. The issue of shock capturing will not be discussed further. However, aliasing errors, how they arise within the DG method and strategies to remove said errors and increase robustness will be discussed at length in this article.

With the above decisions, we arrive at the generic semi-discrete ordinary differential equation (ODE) form of the DG scheme, which can be integrated in time with an appropriate high order explicit or implicit ODE solver, e.g., [50–54].

The resulting DG method is high order accurate and has excellent dispersion and dissipation behavior, e.g., [55, 56]. Furthermore, due to its compact stencil (only interface neighbor data is needed) the DG scheme is well known for its excellent parallel scaling, e.g., [57, 58], and its ability to handle unstructured and non-conforming grids, e.g., [16, 54, 59–63]. These nice properties of the DG methodology are one reason why more and more researchers apply and extend the DG methodology to many different problem setups in computational physics. However, DG is not the perfect discretisation and there are unfortunately some issues that necessitate detailed analysis and discussion.

The remainder of this review article gives the answers to ***why*** we need novel developments, **Section 2**, ***when*** the novel developments started, **Section 3**, ***what*** the key ideas of these novel strategies are, **Section 4**, and ***where*** there are still open questions toward future research directions, **Section 5**.

2 WHY DO WE NEED A NOVEL ROBUST STRATEGY?

Throughout the analysis and discussions in this manuscript we describe different types of stability for a numerical approximation. Principally, we concentrate on the stability and boundedness of the spatial DG discretization.

2.1 On the L_2-Stability of the DG Method

It is easy to show that the DG scheme is L_2-stable for linear advection problems with constant coefficients due to its Galerkin nature, e.g., as a special case in Ref. 64. As a brief illustrative example let's consider the one-dimensional scalar linear advection model

$$u_t + f(u)_x = 0, \tag{11}$$

where $f(u) = au$ with constant velocity a. The respective strong form DG scheme reads

$$\langle U_t, \phi \rangle_E + \left[\left(f^*(U^+, U) - f \right) \phi \right]_{\partial E} + \langle f_x, \phi \rangle_E = 0. \tag{12}$$

We get the discrete evolution of the L_2-norm $\int U^2\, dx$ by inserting $\phi = U$ for the polynomial test function

$$\langle U_t, U \rangle_E + \left[\left(f^*(U^+, U) - f \right) U \right]_{\partial E} + \langle f_x, U \rangle_E = 0. \tag{13}$$

Assuming time continuity, the first term reduces to $\partial_t \int U^2/2\, dx$. We observe that in the volume integral $f_x = aU_x$ is a polynomial of degree $N-1$ and ϕ a polynomial of degree N. Thus, we need the quadrature rule to be exact for polynomials with at least degree $2N-1$. This is guaranteed by all Gauss-Legendre type quadrature rules with at least $N+1$ nodes, such as the Legendre-Gauss-Lobatto (LGL) quadrature.

The volume term contribution is

$$\langle a\, U_x, U \rangle_E = a \left\langle \left(\frac{U^2}{2} \right)_x, 1 \right\rangle_E = \frac{a}{2} \left[U^2 \right]_{\partial E}, \tag{14}$$

which shows that the volume contribution can be *lifted* to the boundary. In total we have

$$\frac{1}{2} \partial_t \int_E U^2 \, \mathrm{d}x + \left[U f^*(U^+, U) - \frac{a}{2} U^2 \right]_{\partial E} = 0. \tag{15}$$

The discrete evolution of the L_2-norm only depends on the choice of the numerical flux function f^*. A simple choice would be the central flux $f^* = \frac{a}{2}(U^+ + U)$. Inserting the central flux into **Eq. 15** gives

$$\frac{1}{2} \partial_t \int_E U^2 \, \mathrm{d}x + \frac{1}{2} [a\, U\, U^+]_{\partial E} = 0. \tag{16}$$

Summing over all elements E in the domain to get the total L_2-norm and assuming periodic boundary conditions, we get

$$\frac{1}{2} \partial_t \| U \|_{L_2}^2 = 0, \tag{17}$$

as $\frac{a}{2} U\, U^+$ is a unique discrete energy flux at every interface and cancels when summing over all elements. Thus, for linear advection, the DG scheme with at least $2N-1$ accurate quadrature is L_2-stable, i.e. the discrete L_2-norm is bounded for all times t. Note, that for quadrature rules with less than $2N$ integration precision, the estimate is not for the exact L_2-norm, but for a discrete L_2-norm corresponding to the quadrature rule chosen.

For linear problems basically all DG variants are stable, but what about nonlinear problems? To address nonlinear problems, there was a crucial contribution by Jiang and Shu in 1994 [65], who demonstrated that for scalar nonlinear hyperbolic problems, the DG method is L_2-stable provided: 1) Exact evaluation of all

integrals are used; 2) Entropy stable numerical surfaces fluxes are used at the element interfaces. The L_2-stability result with conditions 1) and 2) also extends to symmetric variable coefficient hyperbolic systems [66]. Again, as a brief example to illustrate the important steps of the analysis, we consider a scalar one-dimensional problem **Eq. 11** with a simple quadratic flux function $f(u) = \frac{1}{2}u^2$, the so-called Burgers' equation. The DG scheme is, again, given by **Eq. 13** and we get the evolution of the discrete L_2-norm by replacing the test function with the DG solution $\phi = U$. Note, that for this nonlinear problem, the volume integral requires a quadrature rule with higher integration precision to be exact. For the quadratic flux function $f \sim u^2$, the quadrature rule needs $3N-1$ integration precision. With the exact evaluation of the volume integral, its contribution is, once again, lifted onto the boundary of the element to give

$$\left\langle \frac{1}{2}\,(U^2)_x, U \right\rangle_E = \left\langle \left(\frac{U^3}{3}\right)_x, 1 \right\rangle_E = \frac{1}{3}\left[U^3\right]_{\partial E}. \tag{18}$$

The resulting discrete evolution of the L_2-norm is

$$\frac{1}{2}\,\partial_t \int_E U^2\,\mathrm{d}x + \left[U f^*(U^+, U) - \frac{1}{6}\,U^3\right]_{\partial E} = 0, \tag{19}$$

which agin only depends on the choice of the numerical surface flux function f^*. Note, that for the central flux choice $f^*(U^+, U) = \frac{1}{2}(f(U^+) + f(U))$ no stability estimate can be derived, as potentially the L_2-norm could grow without bounds. However, for the particular choice

$$f^*(U^+, U) = \frac{(U^+)^2 + U^+\,U + U^2}{6}, \tag{20}$$

we get

$$\frac{1}{2}\,\partial_t \int_E U^2\,\mathrm{d}x + \frac{1}{2}\left[\frac{U\,(U^+)^2 + U^2\,U^+}{3}\right]_{\partial E} = 0. \tag{21}$$

Following the same arguments as above for the linear advection problem, we sum over all of the elements in the domain and obtain L_2-stability for the nonlinear scalar hyperbolic problem for quadrature rules with at least $3N-1$ integration precision.

Unfortunately, even ignoring the practical issues with the assumption of exact integration for a moment, the results of Jiang and Shu cannot be directly extended to general nonlinear hyperbolic systems, e.g., the compressible Euler equations. A key step in the analysis of Jiang and Shu is that the test functions ϕ in the DG formulation **Eq. 10** are replaced with the discrete DG solution

$$U\left(\vec{x}, t\right) = \sum_{j=0}^{P(N)} U_j(t)\,\phi_j\left(\vec{x}\right), \tag{22}$$

which itself is a linear combination of the test functions $\{\phi_j\}_{j=0}^{P(N)}$ and, hence, an element of the test function space. While this gives an L_2-norm estimate for symmetric systems, this approach does not lead to a proper norm estimate (in the continuous as well as in the discrete case) for general nonlinear systems. This lack of a stability estimate, even when using "exact integration," is the explanation why the DG method can still crash for complex PDEs, e.g., [67].

2.2 On the Entropy Stability of the DG Method

In the previous section it was shown for a scalar nonlinear conservation law that the DG solution U is bounded in the L_2-norm if a special choice of the numerical surface flux is chosen, such as the one in **Eq. 20**. We conjecture that an analogous statement of nonlinear stability should be true for systems of nonlinear conservation laws. However, in general, stability in the L_2-norm is insufficient to exclude unphysical phenomena such as expansion shocks [68]. To remove the possibility of such phenomena we generalize the notion of a stability estimate for nonlinear problems.

Before we discuss the mathematics of a general nonlinear hyperbolic system, a detour is taken to examine an important underlying physical principle. In particular, we introduce concepts from thermodynamics, which is a branch of physics relating the heat, temperature, or entropy of a given physical system to energy and work. The laws of thermodynamics are some of the most fundamental laws in all of physics. This is because they play an important role in describing how, and predicting why, physical systems behave and evolve the way that we observe them. Moreover, thermodynamics provides fundamental rules to decide how a physical system *cannot* behave. That is, what type of solution behavior is physically meaningful and what is not. From a mathematical point of view, we note that satisfying the second law of thermodynamics is not enough to guarantee uniqueness of the PDE solution and that conditions for uniqueness are an active topic of research in the analysis of said PDEs, see for example [69–72].

The first law of thermodynamics concerns the conservation of the total energy in a closed system. The second law of thermodynamics states that the entropy of a closed physical system tends to increase overtime and, importantly, that it cannot shrink. The laws of thermodynamics must be satisfied simultaneously at all times, otherwise a mathematical solution can exhibit strange and obviously incorrect behavior. For example, a fluid that only conserves its total energy but does not take care on the entropy, i.e. satisfying the first law of thermodynamics but not the second law, could transfer all of its internal energy into kinetic energy. The result would be a very fast, but very cold jet of air. Such a flow configuration has never been observed in nature. This discrepancy is removed when incorporating the second law of thermodynamics where the transfer of energies are regulated. For reversible processes the entropy remains constant over time (isentropic) and the time derivative of the total system entropy is zero. For irreversible processes the entropy increases and the time derivative is positive. Solution dynamics where the total system entropy shrinks in time are never observed and deemed unphysical.

To discuss these ideas in a mathematical context consider a system of nonlinear hyperbolic conservations laws

$$\mathbf{u}_t + \vec{\nabla}_x \cdot \overleftrightarrow{\mathbf{f}}(\mathbf{u}) = 0, \tag{23}$$

where we take the viscous flux components in **Eq. 1** to be zero. Typically, the diffusion terms are dissipative in nature and are mostly uncritical. A prototypical example of a purely hyperbolic system of nonlinear conservation laws is the compressible Euler equations modeling inviscid gas dynamics. A smooth solution that satisfies the system of PDEs **Eq. 23** corresponds to a reversible process. One of the difficulties, either analytically or numerically, of nonlinear hyperbolic PDEs is that the solution may develop discontinuities (e.g., shocks) regardless of the continuity of the initial conditions [73]. A discontinuous solution of **Eq. 23** corresponds to an irreversible process and dissipates entropy.

To mathematically account for possible discontinuous solutions, system **Eq. 23** is considered in its weak form. Just as in the Galerkin discretisation in **Section 1**, the weak form of the PDE is found by multiplying the governing equations by a smooth test function $\phi(t, \vec{x})$ with compact support and integrating over $\mathbb{R}^+ \times \mathbb{R}^3$. Integration-by-parts is again applied to move the derivatives onto the test function and weaken the smoothness requirements on possible solutions. Hence, weak solutions of system **Eq. 23** satisfy

$$\int_{\mathbb{R}^+\times\mathbb{R}^3} \mathbf{u}^T \phi_t \, dt \, \mathrm{d}\vec{x} + \int_{\mathbb{R}^+\times\mathbb{R}^3} \overleftrightarrow{\mathbf{f}}^T \vec{\nabla}_x \phi \, \mathrm{d}t \, \mathrm{d}\vec{x} = \int_{\mathbb{R}^3} \mathbf{u}_0^T\left(\vec{x}\right)\phi \, \mathrm{d}\vec{x}. \quad (24)$$

Another form of the conservation law is its integral form that, under the assumption of differentiable fluxes, arises from Gauss' theorem

$$\int_\Omega \mathbf{u}_t \, \mathrm{d}\vec{x} + \oint_{\partial\Omega} \overleftrightarrow{\mathbf{f}}(\mathbf{u}) \cdot \vec{n} \, \mathrm{dS} = 0, \quad (25)$$

and holds for arbitrary control volumes Ω, e.g., [74]. Unfortunately, weak solutions of a PDE are, in general, not unique and must be supplemented with extra admissibility criteria in order to single out the physically relevant solution [75–77]. This is precisely where the laws of thermodynamics play a pivotal role because, as already discussed, due to their intrinsic ability to select physically relevant solutions. In most applications, e.g., compressible fluid dynamics or astrophysics, the total entropy is not part of the state vector of conservative variables **u**. However, we know from the discussion above that for reversible (isentropic) processes the total entropy *is* a conserved quantity. Where is this conservation law "hiding"?

It turns out that there are additional conserved quantities, e.g., the entropy, which are not explicitly built into the nonlinear hyperbolic system **Eq. 23** but are still a consequence of the PDE. In order to reveal this auxiliary conservation law we define a convex (mathematical) entropy function $s = s(\mathbf{u})$ that is a scalar function and depends nonlinearly on the conserved variables **u**. This allows the definition of a new set of *entropy* variables

$$\mathbf{w} = \frac{\partial s}{\partial \mathbf{u}}, \quad (26)$$

that provides a one-to-one mapping between the conservative variable space and entropy space [78]. If we contract the nonlinear hyperbolic system **Eq. 23** from the left with the entropy variables **w** we have

$$\mathbf{w}^T\left(\mathbf{u}_t + \vec{\nabla}_x \cdot \overleftrightarrow{\mathbf{f}}(\mathbf{u})\right) = 0. \quad (27)$$

From the definition of the entropy variables and assuming continuity in time, we know that

$$\mathbf{w}^T \mathbf{u}_t = s_t. \quad (28)$$

Further, each of the flux vectors in the coordinate directions x_i must satisfy a compatibility condition

$$\mathrm{w}^T\left(\mathbf{f}_i\right)_{x_i} = \left(f_i^s\right)_{x_i}, \quad (29)$$

where f_i^s, $i = 1,2,3$ is a corresponding entropy flux [78]. We point out that a chain rule is the linchpin of the manipulations **Eqs. 28** and **29** to move from the space of conservative state variables into the space of entropy variables. In the continuous setting this is not an issue under certain continuity assumptions. However, in a numerical setting it is extraordinarily difficult (or even impossible) to recover the chain rule with discrete differentiation, e.g., [79]. We postpone the discussion on this issue and what it means for a high order DG numerical approximation to **Section 4**.

By definition of a mathematical entropy, contracting the nonlinear hyperbolic system into entropy space, as in **Eq. 27**, and assuming the solution is smooth (i.e. a reversible process) results in the auxiliary conservation law for the entropy

$$s_t + \vec{\nabla}_x \cdot \vec{f}^s = 0. \quad (30)$$

The corresponding integral form of the entropy conservation is given by

$$\int_\Omega s_t \, \mathrm{d}\vec{x} + \oint_{\partial\Omega} \vec{f}^s \cdot \vec{n} \, \mathrm{dS} = 0, \quad (31)$$

for an arbitrary volume Ω. For irreversible processes, physical entropy is increasing. In the mathematical community however, entropy is defined as a decaying function and hence the entropy conservation law **Eq. 31** becomes the entropy inequality [78]

$$\int_\Omega s_t \, \mathrm{d}\vec{x} + \oint_{\partial\Omega} \vec{f}^s \cdot \vec{n} \, \mathrm{dS} \le 0, \quad (32)$$

for discontinuous solutions.

As an illustrative example for the contraction of a nonlinear hyperbolic system of PDEs into entropy space, we consider the compressible Euler equations of gas dynamics in one spatial dimension

$$\begin{bmatrix} \rho \\ \rho v \\ E \end{bmatrix}_t + \begin{bmatrix} \rho v \\ \rho v^2 + p \\ (E+p)v \end{bmatrix}_x = 0, \quad (33)$$

with the ideal gas assumption

$$p = (\gamma - 1)\left(E - \frac{\rho v^2}{2}\right), \quad (34)$$

where γ is the adiabatic constant. The convex entropy function $s(\mathbf{u})$ for the compressible Euler equations is not unique [80]. However, a common choice for the mathematical entropy function is the scaled negative thermodynamic entropy [80–84]

$$s(\mathbf{u}) = -\frac{\rho\varsigma}{\gamma - 1}, \qquad \varsigma = \ln(p) - \gamma\ln(\rho), \tag{35}$$

with the corresponding entropy flux $f^s = v\, s(\mathbf{u})$. From this definition of the mathematical entropy function we get the entropy variables

$$\mathbf{w} = \frac{\partial s}{\partial \mathbf{u}} = \left[\frac{\gamma - \varsigma}{\gamma - 1} - \frac{\rho v^2}{2p}\,,\ \frac{\rho v}{p}\,,\ -\frac{\rho}{p}\right]^T, \tag{36}$$

and see that each component of the entropy variables is a highly nonlinear function of the state vector components. Regardless, the variables **Eq. 36** contract the one dimensional compressible Euler equations into entropy space and, when integrated over the domain Ω, become the entropy conservation law **Eq. 31** for smooth solutions or the entropy inequality **Eq. 32** for discontinuous solutions. It is worth noting that the entropy variables $\mathbf{w}$ are further useful in the analysis of the system, as they allow to derive a symmetric form of the PDE [85].

The discrete equivalent of the entropy inequality **Eq. 32** is referred to as *entropy stability*. It is a generalization of the L_2-stability statement to systems of nonlinear hyperbolic PDEs, e.g., [79, 86]. An additional requirement built into the entropy stability condition is that the fluxes $\mathbf{f}$ remain bounded [68], which restricts the flow to physically realisable states, e.g., positive density and pressure in gas dynamics. Overall, entropy stability ensures that a numerical approximation obeys the fundamental laws of thermodynamics and is viewed as an important quality to capture [86–88]. But it is an active area of research to investigate the role of entropy stability and how it fits into the question of provable nonlinear stability [82, 89, 90].

At present, we restrict ourselves to one of the key ingredients in the analysis to derive a nonlinear entropy stability estimate for general nonlinear hyperbolic systems, see e.g., [78, 86]. It is natural to develop an entropy stable DG approximation because the continuous and discrete analysis both rely on a weak form of the governing equations. However, for entropy stability the nonlinear system is not multiplied by the solution $\mathbf{u}$ as was the case for the L_2-stability analysis. Instead the equation is multiplied with the entropy variables $\mathbf{w}$ **Eq. 26**, which are nonlinear functions of the state $\mathbf{u}$, see e.g., the compressible Euler equations in (**36**). Thus, a direct combination of this approach with the analysis of Jiang and Shu from **Section 2.1** is not possible. For a polynomial DG ansatz $\mathbf{U}$, the discrete entropy variables $\mathbf{W} = \mathbf{W}(\mathbf{U})$ are no longer polynomials of degree N and do not belong to the space of test functions ϕ. Hence, it is not allowed to replace the test functions ϕ in the DG formulation **Eq. 10** with $\mathbf{W}$. Technically, only a projection of $\mathbf{W}$ onto the space of polynomials with degree N can be inserted; however, in this case the analysis does not lead to an entropy stability estimate as the chain rule holds for the full entropy variables and not their projections.

To overcome the issue with the test function space and to enable an entropy stability estimate for the DG method, Hughes et al. [91] as well as Hildebrand and Mishra et al. [92–94] used a space-time DG approach with an ansatz directly written in terms of the entropy variables. The idea is to make the DG ansatz in entropy space, i.e. to approximate the entropy variables

$$\mathbf{w}\left(\mathbf{u}\left(\vec{x}, t\right)\right)\Big|_E \approx \mathbf{W}\left(\vec{x}, t\right) = \sum_{j=0}^{P(N)} \mathbf{W}_j(t)\,\phi_j\left(\vec{x}\right), \tag{37}$$

with a polynomial of degree N. Ignoring the time discretisation for brevity, the DG formulation changes to

$$\begin{aligned}\langle \mathbf{u}(\mathbf{W})_t, \phi\rangle_E + \int_{\partial E} \phi^T \left(\mathbf{f}^*\left(\mathbf{u}(\mathbf{W})^+, \mathbf{u}(\mathbf{W})\right) - \mathbf{f}\left(\mathbf{u}(\mathbf{W})\right)\cdot\vec{n}\right) \mathrm{dS} \\ + \left\langle \vec{\nabla}_x \cdot \overleftrightarrow{\mathbf{f}}\left(\mathbf{u}(\mathbf{W})\right), \phi\right\rangle_E = 0,\end{aligned} \tag{38}$$

which shows that the scheme is still formulated in conservative form, however all the conserved variables $\mathbf{u}$ now depend implicitly on the polynomial approximation of the entropy variables $\mathbf{W}$. As this approach is naturally implicit, a straightforward and elegant extension of the scheme in time is to use a temporal DG scheme on top of the spatial DG scheme, resulting in a fully implicit space-time DG formulation. Hildebrand and Mishra proved that the resulting discretisation is entropy stable provided: i) Exact evaluation of all integrals; ii) Entropy stable numerical fluxes at the spatial surface integrals and upwind fluxes (due to causality) in the temporal surface integrals are used. These conditions on the space-time DG approximation are very similar to those imposed by Jiang and Shu for the scalar nonlinear hyperbolic case discussed for the nonlinear Burgers' equation.

Unfortunately, the assumption 1) on exact integration is extremely difficult to guarantee and implement in practical simulations, which we describe in more detail in the next subsection. Without proper exact quadratures of the integral terms, the chain rule does not hold discretely. Such inexact approximations of functions are referred to as aliasing errors. These can occur in realistic simulations and might cause instabilities. Thus, robustness is still an issue and ad hoc dealiasing or stabilization mechanisms, e.g., artificial diffusion [92] are necessary.

2.3 The Unpleasant Role of Numerical Integration, Nonlinearities, Variational Crimes, and Aliasing

As described above it is necessary to discretize the variational formulation and make certain choices in the DG approximation such as the polynomial basis functions and, especially, the discrete integration to approximate the surface and volume integrals. Unfortunately, only in the rarest and simplest cases is it possible to avoid these discretisation steps and use an exact evaluation of the integrals. Hence, the notion of "variational crimes" is introduced to express the steps necessary to turn

the formulation into an actual algorithm that can be implemented.

One of the biggest problems when discrediting nonlinear advection-diffusion problems is that in many interesting cases, the nonlinearity is non-polynomial. Our exemplary problem, the compressible Euler equations depend on the mass density u_1, the momentum density in the x direction u_2, and energy density u_3. Often, these are also denoted as $u_1 = \rho$ and $u_2 = \rho v$, where v is the velocity. We compute the velocity from the conserved variables as

$$v = \frac{u_2}{u_1}. \tag{39}$$

This is important in the context of the DG discretisation because if the variables u_1 and u_2 are polynomials of degree N, the velocity v is not a polynomial, but a rational function. This occurs not only for the velocity but also for other quantities that are needed to evaluate the advective fluxes $\mathbf{f}$, such as e.g., the pressure p. Hence, the fluxes $\mathbf{f}$ are no longer polynomials of degree N, and possibly rational functions, as in case of the compressible Euler equations. When approximating the integrals with high order quadrature formulae, such as the Legendre-Gauss rules, it is important to realize that these numerical integration rules are constructed for polynomial integrands. Hence, in theory, they cannot integrate non-polynomial functions exactly no matter how many quadrature nodes are considered.

If we focus for instance on the strong form DG volume integral from **Eq. 38**, we see that the core part to evolve the DG solution in time is an L_2 projection of the flux divergence onto the polynomial basis $\langle \overrightarrow{\nabla}_x \cdot \overleftrightarrow{\mathbf{f}}, \phi \rangle_E$. If this projection is not evaluated exactly, due to either the aforementioned variational crimes, the nonlinearities of the flux function, or a combination of the two, the exact L_2 projection turns into a discrete projection, most often taking the form of an interpolation at the quadrature nodes. This is a subtle but important observation. In contrast to an exact L_2 projection, which cleanly "cuts out" high order content of the flux divergence with polynomial degrees larger than N, a *discrete* L_2 projection interprets (i.e. aliases) some of the high order content as part of the projection polynomial. This artificially and unpredictably decreases or increases the polynomial coefficients of the projection. This "incorrect interpretation" of high order content is also well known in Fourier analysis and signal processing. If the sampling rate (in this case the number of Legendre-Gauss quadrature nodes) is not high enough according to the Nyquist theorem, high frequency data (high order content) gets interpreted as low frequency data (onto the polynomial of degree N) and pollutes the result. This analogy to Fourier analysis illustrates the possibility that high frequency information can masquerade as low frequency information when represented on a discrete and unresolved grid. This is the fundamental issue often termed *aliasing*. As the issue of the discrete projection onto a space of polynomials is similar in spirit, the term aliasing is also often used in the DG community, as well as the spectral and finite difference communities, to give potential consequences of the variational crimes a name. In summary, basically all of the DG algorithms for nonlinear advection dominated problems have the issue of inexact evaluation of the integrals and hence all DG algorithms have aliasing errors.

Unfortunately, these aliasing issues are not simply an abstract and "ugly" theoretical oddity without practical consequences. On the contrary, aliasing plays an important role when using DG methods for realistic complex applications to model nonlinear phenomena. It is worth pointing out that one of the advantages of DG, it's very low dissipation errors, are in this particular point of view also it's biggest problem. Due to the inherent low numerical dissipation in a high order DG method, there is no in-built self-defence against the aliasing issues and any instability that they may create. A repercussion of this fact is that it has become naturalized in the numerics community that especially the high order variants of the DG method, with *very* low dissipation errors, have robustness issues in practical applications. For instance DG approximations of the compressible Euler and Navier-Stokes equations are known to sometimes fail due to aliasing instabilities, e.g., [39]. This instability can manifest itself through the observation that the kinetic energy artificially grows in the simulation, while the inner energy decreases. Note, that the total energy is conserved by construction with the DG method; however, this exchange of kinetic and internal energy is unphysical and violates the second law of thermodynamics and is purely a result of the variational crimes (inexact integration).

An obvious solution to these problematic variational crimes, nonlinearities, and aliasing is to decrease their deleterious effects as much as possible. While technically unavoidable in the strictest mathematical sense, it is possible to increase the amount of Legendre-Gauss quadrature nodes to evaluate all integrals "consistently" such that the inexactness errors are on the order of machine precision, see e.g., Kirby et al. [95]. This approach is quite effective and immediately has a positive stabilizing effect on many applications with nonlinear PDEs, see e.g., [39, 96]. However, it is clear that the computational complexity drastically increases when arbitrarily increasing the number of quadrature nodes. Hence, one often tunes the increased number of Legendre-Gauss quadrature nodes and takes as many as needed to make a simulation stable—which is, of course, unsatisfying and ad hoc, as it highly depends on the particular problem setup. In comparison, another approach is to not directly fight the variational crimes themselves, but the consequences they induce. This is achieved by applying well designed filters to the solution, with the purpose to clip out higher order aliasing content in an effort to decrease its effect, e.g., [39]. It is needless to say that this filtering approach is also very ad hoc and depends on many parameters that need tuning depending on the particular problem one wishes to solve with the high order DG scheme.

While these ad hoc stabilization techniques are reinforced by little to no mathematical analysis or rigor, the prevailing consensus in the DG community has been that, in practice, they work reasonably well. In that, consistent integration (sometimes referred to as over-integration) and filtering increase the robustness of high order DG methods to a level such that they could be applied to model challenging physical problems allowing them to shine with their high order accuracy

and low dispersion and dissipation errors. Especially in the turbulence community, many research groups started to apply high order DG methods with stabilization in the context of implicit large eddy simulation with excellent results competitive with others from the broader numerics community, e.g., [97, 98]. However, in Moura et al. [67] it was reported that certain configurations of the DG method for the inviscid Taylor Green vortex problem kept crashing, even when drastically increasing the number of quadrature nodes in the surface and volume integrals. In fact, the amount of quadrature points was increased up to the point where the DG scheme was no longer computationally feasible, but the simulations still crashed. The inviscid Taylor Green vortex setup in this case was used to investigate the case of a very high Reynolds number flow with severe under resolution common in realistic turbulence setups. These findings were also verified by the authors of this review article and have a strong consequence for the DG community. While the ad hoc stabilization techniques were "good enough" in the sense that they helped to make the DG scheme run for a broad range of interesting problems, this approach is apparently not bullet proof. Further, it is impossible to tell a priori for which cases the stabilization will work and for which cases it will not.

This one example, where the high order DG scheme was not stable and could not finish the simulation illustrates, fundamentally, that the removal of aliasing and variational crimes cannot be reliably done in an ad hoc fashion. Instead we need a better understanding of these aliasing errors and how they can be removed from inexact and/or under resolved discretisations. Furthermore, we require a mathematically sound approach to address these aliasing errors in the DG approximation. That is, we need a novel strategy to design robust high order DG methods to approximate the solution of nonlinear advection-diffusion systems.

3 WHEN DID THE NOVEL DEVELOPMENT START?

In 2013, two landmark results completely reshaped the development of the DG method moving forward. First, in his PhD thesis, Fisher extended the work on entropy stable schemes LeFloch and Rhode [99] as well as the high order entropy stable schemes of LeFloch et al. [100] to the summation-by-parts (SBP) finite difference framework with high order boundary closures in Refs. 88 and 101 respectively. LeFloch et al. and Fisher et al. found that the entropy stability estimates for low-order FV methods, developed by Tadmor [78], can be extended to high order accuracy. Whereas the high order reconstruction of LeFloch et al. was for periodic domains (i.e. without considering finite domain boundaries), the SBP finite difference framework includes special boundary closures and are applicable for finite domains. Kreiss et al. [102–105] introduced the SBP finite difference framework to specifically mimic integration-by-parts. Integration-by-parts is a valuable tool for the construction of stability estimates. Further discussion on SBP is given by, e.g., Olsson [106, 107], Strand [108], Nordström [109] and Svärd and Nordström [110].

To briefly introduce the main ideas of the classic SBP finite difference framework, we consider a discretisation in one spatial dimension on a finite interval $E = [-1,1]$. Within this interval, we consider a set of $N+1$ regular grid nodes x_j that include the boundaries $x_0 = 1$ and $x_N = 1$. On this grid, a continuous function $u(x,t)$ is represented as the grid node values $U_j(t) = u(x_j, t)$. In short notation, we collect the nodal values into the vector quantity U. For the approximation of the PDE, we need two discrete operators: One that approximates integration, $\mathcal{M} \in \mathbb{R}^{(N+1)\times(N+1)}$; and one that approximates differentiation, $\mathcal{D} \in \mathbb{R}^{(N+1)\times(N+1)}$. In this article, we only consider diagonal matrices $\mathcal{M}$, sometimes referred to as diagonal norm SBP finite difference operators. With these operators we have

$$\int_E u(x)\, v(x)\, \mathrm{d}x \approx U^T \mathcal{M}\, V \quad \text{and} \quad \frac{d}{dx}\, u(x)|_{x_j} \approx (\mathcal{D}\, U)_j. \tag{40}$$

The discrete integration and differentiation need to be compatible for a SBP operator to satisfy the property

$$(\mathcal{M}\,\mathcal{D}) + (\mathcal{M}\,\mathcal{D})^T = \mathcal{B}, \tag{41}$$

where $\mathcal{B}$ is the boundary integral evaluation operator with $\mathcal{B} = \mathrm{diag}(-1, 0, \ldots, 0, 1)$. Multiplying **Eq. 41** by grid values U^T of an arbitrary function $u(x)$ from the left and the approximation V of an arbitrary function $v(x)$ from the right gives

$$U^T (\mathcal{M}\,\mathcal{D})\, V + U^T (\mathcal{M}\,\mathcal{D})^T V = U^T \mathcal{B}\, V = V_N\, U_N - V_0\, U_0. \tag{42}$$

Grouping terms and using that $\mathcal{M}$ is diagonal such that $\mathcal{M} = \mathcal{M}^T$ we have

$$U^T \mathcal{M}\, (\mathcal{D}\, V) + (\mathcal{D}\, U)^T \mathcal{M}\, V = V_N\, U_N - V_0\, U_0, \tag{43}$$

which is a discrete approximation of the integration-by-parts formula for the corresponding functions u and v

$$\int_E u\, \frac{dv}{dx}\, \mathrm{d}x + \int_E \frac{du}{dx}\, v\, \mathrm{d}x = [u\, v]_{-1}^{1}, \tag{44}$$

hence the name summation-by-parts.

With the SBP property, it is directly possible to show L_2-stability of the finite difference scheme for constant coefficient linear advection problems. Starting again as an example with the scalar problem **Eq. 11** and the linear flux $f = au$, we get the following SBP finite difference semi-discretisation

$$\partial_t U + a\, \mathcal{D}\, U = 0. \tag{45}$$

As stated above, one motivation of the SBP framework is to mimic the energy analysis of finite element discretisations like that found in the DG analysis presented above for linear advection. We proceed and first get a corresponding Galerkin type variational form by multiplying with the discrete integration matrix $\mathcal{M}$

$$\mathcal{M}\, \partial_t U + a\, \mathcal{M}\, \mathcal{D}\, U = 0. \tag{46}$$

This form is valid for all grid functions V^T multiplied from the left and hence is a direct approximation of the variational Galerkin

form, e.g., **Eq. 12**. We point out that this finite difference approximation ignores the surface terms that are specific to the discontinuous Galerkin scheme. Here, the arbitrary grid function V^T takes the role of the test function ϕ. Hence, we can mimic the next step in the analysis, i.e. replacing the test function with the DG solution $\phi = U$, by multiplying **Eq. 46** with U^T from the left

$$U^T \mathcal{M}\, \partial_t U + a\, U^T \mathcal{M}\, \mathcal{D}\, U = 0. \tag{47}$$

The volume term can be reformulated with the SBP property to move the discrete derivative onto the test function and generate boundary data

$$\begin{aligned} U^T \mathcal{M}\, \mathcal{D}\, U &= U^T \left(\mathcal{B} - (\mathcal{M}\, \mathcal{D})^T\right) U = U^T \mathcal{B} U - U^T (\mathcal{M}\, \mathcal{D})^T U \\ &= U^T \mathcal{B} U - U^T (\mathcal{M}\, \mathcal{D}) U. \end{aligned} \tag{48}$$

From this step we see that the contribution of the volume terms in the SBP finite difference scheme can be again lifted to the interval boundaries

$$a\, U^T \mathcal{M}\, \mathcal{D}\, U = \frac{a}{2}\, U^T\, \mathcal{B}\, U, \tag{49}$$

but this time *without* any assumption on a necessary quadrature rule precision. In fact, this is general and holds for all diagonal norm SBP finite difference operators. Again, with the assumption of time continuity and periodic boundary conditions (U_N=U_0), it follows that the discrete L_2-norm of the SBP finite difference solution $\|U\|^2_{SBP} = U^T\, \mathcal{M}\, U$ is bounded for all t.

We emphasize again that neither the reconstruction techniques of LeFloch et al., nor the SBP finite difference framework as a whole, depend on integration or exact evaluation of integrals. Thus, in contrast to the DG stability results discussed above, the stability results obtained for SBP finite differences by Fisher et al. *do not* assume exact evaluation of any integrals. Thus, such methods yield efficient algorithms with feasible implementations that have provable stability estimates.

The second important result was separately discovered in 2013 by Gassner [111]. He realized that the base operators of the nodal discontinuous Galerkin spectral element method (DGSEM) have the diagonal norm SBP property as long as the collocation nodes $\{x_j\}_{j=0}^N$ and weights $\{\omega_j\}_{j=0}^N$ were chosen to be those associated with LGL quadrature. It is interesting to note, that earlier, in 2010, Kopriva and Gassner [34] already found out that for DGSEM with LGL quadrature, the weak DG formulation and the strong DG formulation are discretely equivalent. As shown in **Eqs. 9** and **10**, the weak form and strong form can be transformed into one another with integration-by-parts. Thus, when both forms are discretely equivalent, it basically means that discrete integration-by-parts, i.e., SBP, holds. We point out that in 1996, in the context of spectral methods with Chebyshev-Lobatto nodes or LGL nodes, Carpenter and Gottlieb [112] showed a similar property as SBP for these spectral operators, however they assumed that integration-by-parts holds for the proof. The results in Refs. 34 and 111 complete their findings as they remove the assumption of exact integration.

In the nodal DGSEM-LGL framework, similar to the finite difference framework, the solution coefficients of the DG polynomial are nodal values $U_j(t)$ at the location of the LGL nodes x_j. The nodal DG polynomial is represented with Lagrange basis functions $\{\ell_j(x)\}_{j=0}^N$ spanned with the LGL nodes

$$u(x,t)|_E \approx U(x,t) = \sum_{j=0}^{N} U_j(t)\, \ell_j(x), \tag{50}$$

which have the Kronecker delta property that $\ell_j(x_i) = \delta_{ij}$, i.e., 1 if $i = j$ and 0 otherwise. With this choice of basis function and quadrature rule, it is possible to find discrete versions of the corresponding integral operator and the differentiation operator. For the integral we consider

$$\int_E u(x)\, v(x)\, \mathrm{d}x \approx \sum_{j=0}^{N} \omega_j u\left(x_j\right) v\left(x_j\right) = U^T\, \mathcal{M}\, V, \tag{51}$$

with $\mathcal{M} = \mathrm{diag}(\omega_0, \ldots, \omega_N)$ and the vector of LGL nodal values U and V. Furthermore, we have for the discrete differentiation

$$\frac{d}{dx} u(x,t)\big|_{x_i} \approx U'(x_i,t) = \sum_{j=0}^{N} U_j(t)\, \ell_j'(x_i), \tag{52}$$

where we used the short hand notation for the spatial derivative of the Lagrange basis $\frac{d}{dx}\ell(x) = \ell'(x)$. Introducing the differentiation matrix as

$$\mathcal{D}_{ij} = \ell_j'(x_i), \quad i,j = 0, \ldots, N, \tag{53}$$

we get

$$\frac{d}{dx} u(x,t)|_{x_i} \approx (\mathcal{D}\, U)_j. \tag{54}$$

As was shown in Ref. 111, these two discrete operators are again compatible and provide the SBP property

$$(\mathcal{M}\, \mathcal{D}) + (\mathcal{M}\, \mathcal{D})^T = \mathcal{B}, \tag{55}$$

which means that the DGSEM-LGL operators belong to the class of diagonal norm SBP operators. This simple property of one-dimensional discrete integration-by-parts is the basis for a whole polynomial spectral calculus [113] that includes, for instance, discrete version of Gauss' law on curvilinear grids in three spatial dimensions.

Returning to the discussion on stability, the LGL quadrature rule with N+1 points has an integration precision of 2N−1. Thus, the DGSEM-LGL is stable for scalar linear advection as shown above. However, the DGSEM-LGL is not stable for nonlinear problems, e.g., for the quadratic flux function discussed in **Section 2.1** where an integration precision of 3N−1 is necessary i.e., exact integration of the volume terms. But using the SBP property of the DGSEM-LGL operators, it is possible to apply ideas similar to Fisher et al. and construct a novel DGSEM with LGL quadrature, that is discretely L_2-stable for the nonlinear Burgers' equation, *without* the assumption on exact evaluation of the integrals [111]. These first results have been extended and compounded upon for the compressible Euler equations

[114–118], the shallow water equations [63, 119–121], the compressible Navier-Stokes equations [32, 122, 124], non-conservative multi-phase problems [124], magnetohydrodynamics [125, 126], relativistic Euler [127], relativistic magnetohydrodynamics [128], the Cahn-Hilliard equations [129], incompressible Navier-Stokes (INS) [130], and coupled Cahn-Hilliard and INS [131] among many other complex PDE models and DG discretization types e.g., [132].

4 WHAT IS THE KEY IDEA?

To recap the discussion, there are several obstacles that make it difficult to obtain entropy stability estimates for high order DG methods in the case of a general nonlinear system of hyperbolic PDEs: 1) The assumption of exact evaluation of integrals is unfeasible in practice; 2) We need to contract with entropy variables $\mathbf{w}$ that nonlinearly depend on the conservative quantities $\mathbf{u}$, which means that we need to replace the test functions by a projection (or interpolation) of $\mathbf{w(u)}$; 3) We need to satisfy a discrete version of the chain rule to contract the flux divergence into entropy space, i.e. a discrete version of $\mathbf{w}^T (\mathbf{f}_i)_{x_i} = (f_i^s)_{x_i}$, i = 1,2,3, where now the entropy variables and the flux functions are discrete projections and the derivative is replaced with our discrete derivative operator.

In what follows, the key ideas to resolve all three issues are presented. We focus on issue 1) and consider a scalar nonlinear problem with quadratic fluxes discussed above first. Then we ramp-up the complexity in the second subsection and discuss how to extend the novel approach to general systems and how to resolve all issues (i)-(iii).

4.1 On the Conservative Form, Split Forms and Skew-Symmetry

To illustrate the general idea of a split form and how to incorporate it into a high order DG approximation we examine our simple scalar nonlinear hyperbolic conservation law, the Burgers' equation. We start with the conservative form

$$u_t + \left(\frac{u^2}{2}\right)_x = 0, \tag{56}$$

that can be rewritten into its advective form

$$u_t + u\, u_x = 0, \tag{57}$$

which is equivalent in the continuous case for smooth solutions. We can also consider an equivalent combination of the two forms

$$u_t + \alpha \left(\frac{u^2}{2}\right)_x + (1-\alpha)\, u\, u_x = 0, \tag{58}$$

where $\alpha \in \mathbb{R}$ is an arbitrary parameter. This form is called the split form of Burgers' equation, with α being the split form parameter.

While in the continuous case with smooth solutions all of these forms are equivalent, it is important to note that in the discrete case this is not true. Considering the DGSEM operators with N+1 LGL nodes, the volume terms for the conservative form are given by

$$\left(\frac{u^2}{2}\right)_x \approx \frac{1}{2} \mathcal{D}\, \underline{U}\, U, \tag{59}$$

where U is the vector of values of u at the LGL nodes, $\mathcal{D}$ is the DGSEM-LGL derivative operator and $\underline{U} = \mathrm{diag}(U_0, \ldots, U_N)$ is a matrix that has the nodal values U injected onto its diagonal. Analogously, the volume terms for the advective form are

$$u\, u_x \approx \underline{U}\, \mathcal{D}\, U. \tag{60}$$

Only for polynomial functions u with degree $\leq N/2$ and their corresponding nodal values U, do we have

$$\frac{1}{2} \mathcal{D}\, \underline{U}\, U = \underline{U}\, \mathcal{D}\, U. \tag{61}$$

In all other cases, i.e. in the general case for arbitrary nodal vectors U, we get

$$\frac{1}{2} \mathcal{D}\, \underline{U}\, U \neq \underline{U}\, \mathcal{D}\, U. \tag{62}$$

The discrete forms are different because of different aliasing errors. Whereas the conservative form computes a discrete derivative of U^2, the second form computes a "clean" derivative of U, but on the other hand needs to compute the product of two functions U and $\mathcal{D}U$ on a grid with only N+1 nodes. An interesting question is, if we can make use of the different (aliasing) errors in the two forms and find combinations via the split formulation where these errors cancel.

We note that the idea of split formulations was already introduced in the spectral community to develop stable numerical methods for the incompressible Navier-Stokes equations e.g., [133], but is especially prominent in the finite difference fluid dynamics community e.g., [134–139]. Split formulations are used as a built-in dealiasing mechanism to stabilize numerical methods e.g., [140]. Combinations of different forms of the advective terms of the compressible Euler equations yield finite difference approximations that are more robust than the standard conservative ones. In a perfect world, it would be desirable if we could choose the split form parameter α such that the different aliasing errors cancel exactly. Unfortunately, in general, it is not possible to cancel the aliasing errors for each grid node; however, what we will show next is that it *is* possible to cancel the aliasing errors in a way to get a global L_2-stability estimate similar to the estimates by Jiang and Shu [65], but without the assumption of exact integration.

First, we derive the strong DG formulation of the split form of Burgers' **Eq. 58** by multiplying with a test function ϕ, integrating over a grid cell E, inserting a numerical flux function f^* at the grid cell interface to account for the discontinuous nature of our ansatz, and use integration-by-parts again to arrive at

$$\int_E U_t\, \phi \, \mathrm{d}x + \Big[\left(f^*(U^+, U) - f(U)\right) \phi \Big]_{\partial E} + \int_E \Big[\alpha \left(\frac{U^2}{2}\right)_x + (1-\alpha)\, U\, U_x \Big] \phi \, \mathrm{d}x = 0, \tag{63}$$

where the flux is $f(U) = U^2/2$. We consider specifically the DGSEM-LGL variant to get discrete operators that satisfy the

SBP property with diagonal norm (mass matrix) $\mathcal{M}$. We arrive at the DGSEM-LGL variant when we replace the integrals by discrete quadrature with N+1 LGL nodes and when using the same N+1 LGL nodes to span the Lagrange basis functions used for our polynomial ansatz. This gives the following discrete DGSEM-LGL split form

$$\partial_t U + \mathcal{M}^{-1}\mathcal{B}\left[F^* - F\right] + \alpha\frac{1}{2}\mathcal{D}\,\underline{U}\,U + (1-\alpha)\,\underline{U}\,\mathcal{D}\,U = 0, \quad (64)$$

where $\mathcal{B}$ is the boundary evaluation matrix from the SBP property **Eq. 41**, $F = F(U)$ is the vector of collocated nodal flux values i.e., $F_j = f(U_j)\,\forall j$, and F^* is a vector that contains the numerical fluxes at the interfaces "left" and "right" in its first and last entry and is zero elsewhere. The value U^+, again, indicates that the numerical flux functions depends not only on local element values U, but also on the values from the neighbor grid cells. We refer to Gassner [111] for a detailed derivation of this form and its connection to the SBP framework with simultaneous-approximation-terms (SAT).

Next, we follow the standard procedure to derive an L_2-stability estimate by multiplying the scheme with the DG solution and use the quadrature rule to numerically integrate over the element i.e., we multiply by $U^T\mathcal{M}$

$$\begin{aligned} U^T\mathcal{M}\,\partial_t U + U^T\mathcal{B}\left[F^* - F\right] + \alpha\frac{1}{2}U^T\mathcal{M}\mathcal{D}\,\underline{U}\,U \\ + (1-\alpha)\,U^T\,\underline{U}\,\mathcal{M}\mathcal{D}\,U = 0, \end{aligned} \quad (65)$$

where we used the fact that $\mathcal{M}\,\underline{U} = \underline{U}\,\mathcal{M}$ because both matrices are diagonal. Again, we consider the semi-discrete version and assume continuity in time to have

$$U^T\mathcal{M}\,\partial_t U = \frac{1}{2}\partial_t U^2_{\mathcal{M},E}, \quad (66)$$

the evolution of the discrete L_2-norm in the grid cell E. Next, we focus on the volume terms. Note, that in the analysis of Jiang and Shu exact integration was assumed to contract the volume contribution to the surface. This is very important, as it allows direct control over the stability of the scheme with the choice of the numerical interface flux F^*. Without the assumption of exact integration however, we look at the influence of the choice of the split form parameter α instead. We realize that the second term in the volume integral can be transposed $U^T\,\underline{U}\,\mathcal{M}\mathcal{D}\,U = U^T(\mathcal{M}\mathcal{D})^T\,\underline{U}\,U$ and is similar to the first term in the volume integral, except for $(\mathcal{M}\mathcal{D})^T$ is now transposed. Using the SBP property $(\mathcal{M}\mathcal{D})^T = \mathcal{B} - \mathcal{M}\mathcal{D}$ we get

$$\begin{aligned} &\alpha\frac{1}{2}U^T\mathcal{M}\mathcal{D}\,\underline{U}\,U + (1-\alpha)\,U^T\,\underline{U}\,\mathcal{M}\mathcal{D}\,U \\ &\quad = \alpha\frac{1}{2}U^T\mathcal{M}\mathcal{D}\,\underline{U}\,U + (1-\alpha)U^T\,(\mathcal{M}\mathcal{D})^T\,\underline{U}\,U, \\ &\quad = \left(\alpha\frac{3}{2} - 1\right)U^T\mathcal{M}\mathcal{D}\,\underline{U}\,U + (1-\alpha)U^T\,\mathcal{B}\,\underline{U}\,U. \end{aligned} \quad (67)$$

The term with the boundary evaluation matrix $\mathcal{B}$ is a surface term, however the remainder term is a volume term that can either increase or decrease the L_2-norm. Hence, this term can be potentially critical in cases where it increases the norm, as this is an unstable behavior that could lead to break down of the simulation. We note that this volume term is another expression of the aliasing issues. To guarantee that this term does not affect stability, we need to guarantee that it vanishes. We see that there is a single (unique) choice of the split form parameter $\alpha = 2/3$ that cancels the remaining volume term. With this choice, the discrete change of the L_2-norm reads as

$$\begin{aligned} &\frac{1}{2}\partial_t U^2_{\mathcal{M},E} + U^T\mathcal{B}\left[F^* - F\right] + \frac{2}{3}U^T\,\mathcal{B}\,F \\ &\quad = \frac{1}{2}\partial_t||U||^2_{\mathcal{M},E} + U^T\mathcal{B}\left[F^* - \frac{1}{3}F\right] = 0. \end{aligned} \quad (68)$$

This estimate is now analogous to the one with exact integration **Eq. 19** and hence, with the same arguments, the choice of the numerical flux functions as

$$f^{*,\mathrm{EC}}(U^+, U) = \frac{1}{6}\Big((U^+)^2 + U^+\,U + U^2\Big), \quad (69)$$

gives again a discrete stability estimate

$$\partial_t\|U\|^2_{\mathcal{M}} = 0, \quad (70)$$

for the split form DGSEM-LGL with $\alpha = 2/3$ when summing over all grid cells with periodic boundary conditions. It is important to observe that this estimate is discrete in the sense that it did not assume exact integration and that it can be only derived with the particular choice $\alpha = 2/3$ to cancel out the volume contribution of the aliasing errors. With this, we have a novel method where we have solved issue i) mentioned in the beginning of the section.

We note that this particular choice of numerical flux function exactly preserves the discrete L_2-norm. If one considers non-smooth solutions this choice would be inappropriate as for e.g., shocks, because the L_2-norm needs to decrease as u^2 is a mathematical entropy for Burgers' equation. For scalar equations, $s(u) = u^2/2$ is the square entropy (which leads to an L_2-stability estimate e.g., [99, 141]) that gives the simple entropy variables $w(u) = \frac{ds}{du} = u$. Hence, the specific choice of numerical flux **Eq. 72** is often referred to as an entropy conservative (EC) flux function. For an entropy dissipative flux function, there are many choices available. It can be shown that the class of E-fluxes, e.g., [33], are guaranteed dissipative and lead to the estimate

$$\partial_t\|U\|^2_{\mathcal{M}} \le 0. \quad (71)$$

As an example, a simple choice of an entropy dissipative numerical flux function is that of Rusanov

$$\begin{aligned} f^*(U^+, U) &= \frac{1}{2}\left(F(U^+) + F(U)\right) - \frac{1}{2}\,\lambda_{max}\,(U^+ - U), \\ \lambda_{max} &= \max_{U^+,U}\left(\frac{\partial f}{\partial u}\right). \end{aligned} \quad (72)$$

An important question is how we can extend this approach to general nonlinear systems. Following the same ideas, it was possible to derive split forms for the shallow water equations [63, 119, 142], a simplified version of the compressible Euler equations. There are many split forms for compressible Euler

that, for instance, give kinetic energy preserving properties e.g., [143, 144]. However, up to now, no split form for the compressible Euler equations is known that gives the desired discrete entropy stability estimate. The problems are issues ii) and iii) mentioned in the beginning of the section, where we need the discrete chain rule property to contract the volume terms to the surface. Concluding this subsection we revisit the derivations of Burgers' equation and make two important observations.

First, with the proper choice of $\alpha = 2/3$, we get the so-called skew-symmetric form

$$u_t + \frac{1}{3}\left(\left(u^2\right)_x + u\, u_x \right) = 0. \tag{73}$$

Skew-symmetry is strongly connected to entropy, see e.g., Tadmor [85]. Multiplying the spatial derivative term by u as in the L_2-stability analysis gives

$$u\left(\left(u^2\right)_x + u\, u_x \right) = u\left(u^2\right)_x + u^2\,(u)_x = \left(u^3\right)_x, \tag{74}$$

which shows that the skew-symmetric form gives a product-rule type form in the stability analysis that can directly be contracted to the divergence form i.e., contracts to the surface when integrating. In fact, for this simple problem, the chain rule needed for contraction reduces to the simpler product rule. Analogously, we get for the discrete skew-symmetric volume terms of the DGSEM-LGL

$$U^T \mathcal{M}\mathcal{D}\,\underline{U}\,U + U^T\,\underline{U}\,\mathcal{M}\mathcal{D}\,U = U^T \mathcal{B}\,\underline{U}\,U. \tag{75}$$

Thus, in our derivation, we already used a specific discrete version of the chain rule (product rule) to get our estimate. The question is, how to extend this idea to the general case?

The second important observation pioneered for SBP schemes by Fisher in his PhD thesis [88] (in the spirit of earlier work by LeFloch et al. [100]) is that the particular skew-symmetric volume terms $\frac{1}{3}\left[\mathcal{D}\,\underline{U}\,U + \underline{U}\,\mathcal{D}\,U\right]$ can be rewritten for any diagonal norm SBP operator (hence, also for the DGSEM-LGL case) into

$$\begin{aligned}\frac{1}{3}\left[\mathcal{D}\,\underline{U}\,U + \underline{U}\,\mathcal{D}\,U\right]_i &= 2\sum_{j=0}^{N}\mathcal{D}_{ij}\frac{1}{6}\left(U_j^2 + U_j\,U_i + U_i^2\right)\\ &= 2\sum_{j=0}^{N}\mathcal{D}_{ij} f^{*,\mathrm{EC}}\left(U_j, U_i\right)\\ &= \mathbb{D}^{EC} f,\end{aligned} \tag{76}$$

with $f^{*,\mathrm{EC}}$ being the particular numerical flux **Eq. 69** that is symmetric in its arguments and that leads to exact conservation of U^2. We further introduced the shorthand notation $\mathbb{D}^{EC} f$ for the volume term, that indicates that we use a specific derivative operator built on the EC-flux. As a remark, we note that this relation is easy to prove, as the discrete derivative of a constant is zero and then, for instance,

$$\sum_{j=0}^{N}\mathcal{D}_{ij}U_i^2 = U_i^2\sum_{j=0}^{N}\mathcal{D}_{ij} = 0. \tag{77}$$

In combination with the first observation we get the property that this new discrete derivative operator (or divergence operator in the multi-dimensional case) satisfies

$$U^T\,\mathcal{M}\mathbb{D}^{EC} f = \mathcal{B}\,F^s, \tag{78}$$

where F^s is the collocated nodal vector of the entropy flux $f^s = u^3/3$ for Burgers' equation with the square entropy $s(u) = u^2/2$. This relation is the important discrete analogue of the chain-rule property $u f_x = f_x^s$, as it follows with integration that (**82**) is the discrete analogue of

$$\int_Q u f_x\,\mathrm{d}x = \left[f^s\right]_{\partial Q}. \tag{79}$$

We will see in the next subsection, how these observations guide the path to discrete stability estimates for general hyperbolic PDE systems.

4.2 On the Discrete Entropy Stability of the DGSEM-LGL

We have demonstrated how to build a high order skew-symmetric DG approximation of the scalar nonlinear Burgers' equation. To do so required a very particular discrete derivative operator **Eq. 78** that was the key to restore discrete entropy stability. We now discuss how to extend the split form approach to general systems of nonlinear hyperbolic conservation laws. For general nonlinear systems, it is unclear how to explicitly construct the split form to obtain a discrete chain rule property. In particular, the compatibility condition on the physical fluxes obtained when one contracts into entropy space **Eq. 29** that we reproduce here, assuming one spatial dimension, due to their pertinence in the present discussion

$$\mathbf{w}^T\mathbf{f}_x = f_x^s. \tag{80}$$

As previously indicated, the chain rule is either unfeasible or even impossible to directly recover with discrete differentiation. With this in mind we apply the product rule to this compatibility condition between the physical flux $\mathbf{f}$ and the entropy flux f^s to find

$$\mathbf{w}_x^T\mathbf{f} = \left(\mathbf{w}^T\mathbf{f}\right)_x - f_x^s = \left(\mathbf{w}^T\mathbf{f} - f^s\right)_x. \tag{81}$$

A principle motivation for this manipulation is because it is far easier to recover the product rule discretely than it is the chain rule. That is, we already have a particular discrete equivalent for the product rule if the discrete derivative matrix $\mathcal{D}$ is a SBP operator.

Next, we aim to find a discrete version of the new compatibility condition **Eq. 81** following the ideas of Tadmor [78]. Tadmor analyzed low order FV schemes and developed conditions on the numerical surface flux to derive a discretely entropy conserving scheme. In the context of the low order FV methodology, our unknowns in the elements are mean values that are naturally discontinuous across grid cell interfaces. As mentioned above, the idea to resolve these discontinuities with numerical flux functions was also used in the construction of the DG approximation. Consider the contribution to the compatibility condition **Eq. 81** at an arbitrary interface. It depends on the discrete values in the current cell and the direct neighbor of that cell, denoted again

with a "+". We approximate all derivatives with first order differences and define Tadmor's *entropy conservation* condition on the numerical surface flux function

$$\left(\frac{(\mathbf{w}(\mathbf{U}^+)) - \mathbf{w}(\mathbf{U})}{\Delta x}\right)^T \mathbf{f}^{*,\mathrm{EC}}(U^+, U) = \frac{\left(\mathbf{w}(\mathbf{U}^+)^T\mathbf{f}(\mathbf{U}^+) - \mathbf{f}^s(\mathbf{U}^+)\right) - \left(\mathbf{w}(\mathbf{U})^T\mathbf{f}(\mathbf{U}) - f^s(\mathbf{U})\right)}{\Delta x}, \quad (82)$$

where Δx is the size of each grid cell. Equivalently, we arrive at the following general condition on the numerical surface flux for entropy conservation

$$((\mathbf{w}(\mathbf{U}^+)) - \mathbf{w}(\mathbf{U}))^T\mathbf{f}^{*,\mathrm{EC}}(\mathbf{U}^+, \mathbf{U}) = \left(\mathbf{w}(\mathbf{U}^+)^T\mathbf{f}(\mathbf{U}^+) - \mathbf{f}^s(\mathbf{U}^+)\right) - \left(\mathbf{w}(\mathbf{U})^T\mathbf{f}(\mathbf{U}) - f^s(\mathbf{U})\right). \quad (83)$$

For scalar nonlinear problems this condition can be solved explicitly [145]. For example, in the case of Burgers' equation, we have $w(u) = u, f(u) = u^2/2$, and $f^s(u) = u^3/3$ such that solving **Eq. 83**

$$f^{*,\mathrm{EC}}(U^+, U) = \frac{1}{6}\frac{(U^+)^3 - U^3}{U^+ - U} = \frac{1}{6}\left((U^+)^2 + U^+U + U^2\right), \quad (84)$$

which matches the particular entropy conservative flux derived in **Section 4.1**. We note again that the entropy conservative flux is symmetric in its arguments U^+ and U, and is consistent to the physical flux in the sense that for the same arguments we recover the PDE flux $f^{*,\mathrm{EC}}(U, U) = f(U)$.

However, for systems of nonlinear hyperbolic conservation laws **Eq. 83** is a single algebraic condition for a system vector of unknown flux quantities. Therefore, care must be taken to define an entropy conservative numerical flux function that remains physically consistent. That being said, the entropy conservation condition on the numerical surface flux **Eq. 83** is an incredibly powerful statement. Provided we know an explicit form of the entropy variables, the physical flux, and the entropy flux we can define an appropriate numerical flux that ensures entropy consistency for a low order FV numerical approximation. A general form for such a numerical flux was developed by Tadmor [77] defined as a phase integral

$$\mathbf{f}^{*,\mathrm{EC}}(\mathbf{U}^+, \mathbf{U}) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathbf{f}\left(\tilde{W}(\mathbf{U}(\xi))\right) \mathrm{d}\xi, \qquad \tilde{W}(\mathbf{U}(\xi)) = \frac{1}{2}(\mathbf{W}(\mathbf{U}^+) + \mathbf{W}(\mathbf{U})) + \xi(\mathbf{W}(\mathbf{U}^+) - \mathbf{W}(\mathbf{U})). \quad (85)$$

To evaluate the phase integral form of the numerical flux function requires a certain quadrature rule that defines a path through phase space. Though theoretically useful, this phase integral form is computationally prohibitive for practical simulations, even for low order numerical approximations. However, over the past 20 years "affordable" versions of the entropy conservative flux function $\mathbf{f}^{*,\mathrm{EC}}(\mathbf{U}^+, \mathbf{U})$ have been developed for a variety of nonlinear systems like the shallow water equations [119, 146], compressible Euler [81, 83], and ideal magnetohydrodynamics [147].

The key to these numerically tractable versions of the entropy conservative numerical flux function is to evaluate the components of the physical flux at various mean states between U^+ and U. Note that these mean states can take on incredibly complex forms that depend on the arithmetic mean, the product of arithmetic means, or more uncommon quantities like the logarithmic mean. Complete details on the derivation of such numerical flux functions can be found in e.g., [81, 83, 146, 147].

For illustrative purposes we summarize the specific form of an entropy conservative numerical flux for the one dimensional compressible Euler equations with the ideal gas assumption due to Chandrashekar [81]. First, we introduce notation for the arithmetic mean and the logarithmic mean for two quantities a and a^+:

$$\{\{a\}\} = \frac{1}{2}(a^+ + a), \quad a^{\ln} = \frac{a^+ - a}{\ln(a^+) - \ln(a)}. \quad (86)$$

Also, we introduce a variable proportional to the inverse of the temperature

$$\beta = \frac{p}{2\rho}, \quad (87)$$

which simplifies the form of the entropy variables **Eq. 36** to be

$$\mathbf{w} = \left[\frac{\gamma - s}{\gamma - 1} - \beta v^2,\ 2\beta v,\ -2\beta\right]^T. \quad (88)$$

Then, applying the entropy conservation condition **Eq. 83** and many algebraic manipulations later, we arrive at an analytical expression of an entropy conservative numerical flux for the compressible Euler equations

$$\mathbf{f}^{*,\mathrm{EC}}(\mathbf{U}^+, \mathbf{U}) = \begin{bmatrix} \rho^{\ln}\{\{v\}\} \\ \rho^{\ln}\{\{v\}\}^2 + \frac{\{\{\rho\}\}}{2\{\{\beta\}\}} \\ \frac{\{\{v\}\}}{2}\left(\frac{\rho^{\ln}}{\beta^{\ln}(\gamma - 1)} + \frac{\{\{\rho\}\}}{\{\{\beta\}\}}\right) + \frac{\rho^{\ln}\{\{v\}\}}{2}\left(2\{\{v\}\}^2 - \{\{v^2\}\}\right) \end{bmatrix}. \quad (89)$$

So far, the discussion on entropy conservative numerical approximations has all been in the context of low order finite volume methods. It is possible to create a high order entropy aware scheme with ENO [148] or WENO type reconstructions [149]. However, as mentioned above, a critical and remarkable result of Fisher's work is that a low order finite volume entropy conservative scheme can be extended to an arbitrarily high order accurate spatial scheme, when it is based on diagonal norm SBP operators [101]. As was described for Burgers' equation in the last two observations in **Section 4.1**, the crucial part is to move the contribution to the entropy production from the volume terms to the surface via a discrete version of the chain rule. Once stability is only governed by surface contributions, it is possible to control the stability with a proper choice of numerical interface flux

function. In this sense, the analysis of the high order scheme reduces to a similar problem as for the low order finite volume method **Eq. 83** and the well-established theoretical analysis tools and results can be reused. It is worth mentioning that $\mathbf{f}^{*,\mathrm{EC}}$ is not unique and that a particular choice of the entropy conservative flux generates a different split form of the governing equations, e.g., [144, 150, 151].

We return to a strong form DG approximation of a nonlinear hyperbolic system of conservation laws that takes a form identical to **Eq. 38** but now only considered in one spatial dimension

$$\langle \mathbf{U}_t, \phi \rangle_E + \left[\phi^T \left(\mathbf{f}^* \left(\mathbf{U}^+, \mathbf{U}\right) - \mathbf{f}\left(U\right)\right)\right]_{-1}^{1} + \langle \mathbf{f}_x, \phi \rangle_E = 0. \tag{90}$$

Here the standard tools of a nodal DG approximation have been applied:

1. The solution and physical fluxes are approximated with polynomials.
2. Any integrals in the variational formulation are approximated with high order LGL quadrature.
3. The interpolation and quadrature nodes are collocated.

Importantly, these steps mean that the discrete DG differentiation matrix is a SBP operator. Furthermore, we use the entropy conservative numerical surface flux at the interface and the particular discrete derivative projection $\mathbb{D}^{EC}$ defined in **Eq. 76** in the volume contribution to arrive at the entropy conservative DG approximation

$$\phi^T \left(\mathcal{M}\mathbf{U}_t + \left[\mathbf{f}^{*,\mathrm{EC}}\left(\mathbf{U}^+, \mathbf{U}\right) - \mathbf{f}\right]_{-1}^{1} + \mathcal{M}\mathbb{D}^{EC}\mathbf{f} \right) = 0, \quad \forall \phi \in \mathbb{R}^{N+1}. \tag{91}$$

Now, if we take the test function $\phi = \mathbf{W}$ with $\mathbf{W}_j = \mathbf{w}(\mathbf{U}_j)$ evaluated at each LGL node x_j, we obtain

$$\boldsymbol{W}^T \mathcal{M}\mathbf{U}_t = \sum_{j=0}^{N} \omega_j \mathbf{W}_j^T \left(\mathbf{U}_t\right)_j = \sum_{j=0}^{N} \omega_j \left(S_t\right)_j = \langle S_t, 1 \rangle_E, \tag{92}$$

assuming continuity in time. Also, the discrete differentiation operator $\mathbb{D}^{EC}$ moves volume information onto the boundary, see Refs. 32, 101, and 125 for complete details, such that

$$\boldsymbol{W}^T \mathcal{M}\mathbb{D}^{EC}\mathbf{f} = \mathcal{B}F^s. \tag{93}$$

We note that this remarkable property holds for general nonlinear systems with available entropy estimate and a corresponding low order entropy conserving flux $\mathbf{f}^{*,\mathrm{EC}}$. Combining this with the definition of the entropy conservative flux **Eq. 83**, the discrete entropy evolution of the DGSEM-LGL becomes

$$\langle S_t, 1 \rangle_E + \left[F^s\right]_{-1}^{1} = 0, \tag{94}$$

which is the discrete analogue of the integral form of the entropy conservation law discussed in **Section 2.2**. The resulting DGSEM-LGL is entropy conservative by construction and it is important to note we have assumed ***no exactness*** on the integration. From this baseline entropy conservative numerical scheme, that does not dissipate entropy by construction, we can create a high order DGSEM-LGL that enforces the entropy inequality **Eq. 32**. We do so by introducing dissipation at the element interfaces via the choice of the numerical surface flux function, e.g., the Rusanov flux **Eq. 72**. More complex dissipation techniques are also available that dissipate solution information according to the different wave strengths with complete details found in, e.g., [152–156]. We finally note that this discussion was restricted to one spatial dimension for the sake of convenience and simplicity. Extensions to general three dimensional curvilinear coordinate systems are available, see e.g., [32, 101, 116, 125, 157, 158] for details.

4.3 Validation of Robustness and Application to Space Physics of the Entropy Stable DGSEM

In this subsection, we demonstrate two exemplary simulation results of a DG scheme based on the key ideas outlined above. The general split form DGSEM with LGL nodes on three dimensional curvilinear hexahedral unstructured meshes is implemented in the open source software FLUXO (project-fluxo/fluxo at github), written in modern Fortran with a special emphasis on massively parallel CPU based hardware. The main focus of the software is on compressible Navier-Stokes and visco-resistive MHD equations. Time integration of the semi-discrete form is done with a fourth order accurate low storage Runge-Kutta method of Carpenter and Kennedy [159].

For the **validation** of the robustness we revisit an important numerical contribution of Moura et al. [67]. They were the first to report of a test case that the DG scheme with (numerical) exact integration was not able to run, demonstrating, that further improvement on the robustness of the DG methodology was necessary. For this validation test, we consider the compressible Navier-Stokes equations (viscous case) or the compressible Euler equations (inviscid case, basically setting the viscosity parameter to zero). The considered problem is the Taylor-Green vortex in a fully periodic domain, which serves as a test case for a fully periodic turbulent box $[0,1]^3$, that starts with a smooth initial velocity field

$$\begin{aligned} v_1 &= v_0 \sin(2\pi x_1)\cos(2\pi x_2)\cos(2\pi x_3), \\ v_2 &= -v_0 \cos(2\pi x_1)\sin(2\pi x_2)\cos(2\pi x_3), \\ v_3 &= 0, \end{aligned} \tag{95}$$

and transitions to turbulence during its temporal evolution until it reaches a state similar to homogeneous turbulence. The initial density is uniform $\rho = 1$ and the initial pressure is give by $p = p_0 + \frac{1}{16}\left(\cos(4\pi x_1) + \cos(4\pi x_2)\right)\left(\cos(4\pi x_3) + 2\right)$, where p_0 is a background pressure and v_0 the velocity amplitude used to set the initial Mach number. In our case, we choose the Mach number to be $Ma = 0.1$. Unresolved vortical driven flows are especially prone to the aliasing issues discussed above. The difficulty of this test case lies in its wide range of scales when the Reynolds number increases (i.e., for low viscosities), e.g., [39]. For the DGSEM discretisation, we choose the polynomial degree N and the number of grid cells N_Q^3. Thus, the total number of degrees of freedom (DOF) for one conserved quantity is

$(N+1)^3 N_Q^3$. For the numerical flux function at the surface, we use the Rusanov flux.

For the robustness investigation, we consider an inviscid flow (with the viscosity parameter is zero) and focus on three particular setups, where N = 1,3,7 with number of elements N_Q = 56,28,14 respectively. This ensures that for all three computations the overall number of DOF per conserved quantity is equal, about 1.4 million. There are many more investigations of different configurations presented in Ref. 160, but they all demonstrate the same behavior: while the low order variants N = 1,3 seem to be relatively robust with full integration, the higher the polynomial degree, the less stable the DG method becomes. And for the case N = 7 with N_Q = 14 the simulation crashed at about a simulation time of t_{crash} = 8.4, even when increasing the quadrature nodes from 8^3 up to 32^3 = 32768 per element. In contrast, the novel entropy stable DGSEM with standard LGL nodes runs all configurations without crashing.

Furthermore, it is possible to run this challenging test case even without any artificial dissipation, i.e. with the $F^{*,\mathrm{EC}}$ numerical flux instead of the Rusanov flux function. This is very interesting, as it allows us to fully observe and control the artificial numerical dissipation generated by the scheme. To demonstrate this, we consider again the Taylor-Green vortex test case, but this time with non-zero viscosity such that the Reynolds number is Re = 1600. In this Navier-Stokes case, it is possible to relate the kinetic energy decay over time with the temporal behavior of the enstrophy to get an estimate for the Reynolds number. In theory, this should be Re = 1600 for the simulation. In practice, the finite resolution causes numerical errors such as dispersion and dissipation, e.g., [55]. We present two results in **Figure 1** for the viscous test case with N = 7 and N_Q = 8.

We note, that this test case would crash for the standard DG scheme, however for the presented novel DGSEM-LGL with the proper discrete chain rule, it runs with the dissipative Rusanov numerical flux $F^{*,\mathrm{Rusanov}}$ (entropy stable DGSEM-LGL) and even with the non-dissipative numerical flux $F^{*,\mathrm{EC}}$ (entropy conservative DGSEM-LGL). After an initial transition zone, the entropy conservative scheme retains the physical Reynolds number remarkably well with $Re_{numerical}$ = 1600 and the simulation is virtually dissipation free throughout the temporal evolution. The entropy stable variant clearly introduces stabilizing dissipation as soon as the spatial scales can no longer be resolved. It is interesting to note, that these results hint toward the possibility of quantifying and controlling the artificial dissipation of the DGSEM for under resolved turbulence and use this to construct high fidelity turbulence models, see e.g., for a proof of concept [161].

For an exemplary **application** we consider a complex test case from space physics. We focus on the electrodynamic and plasma interaction of the moon Io with the strong magnetic field of Jupiter. Io is embedded in a dense plasma torus, induced by the magnetosphere of Jupiter, and it exhibits interesting plasma flow characteristics containing steep gradients and discontinuities [162]. The general problem setup is illustrated in **Figure 2**. Neglecting neutral density, relativistic, viscous, resistive, and Hall effects, this MHD flow within Io's plasma torus can be

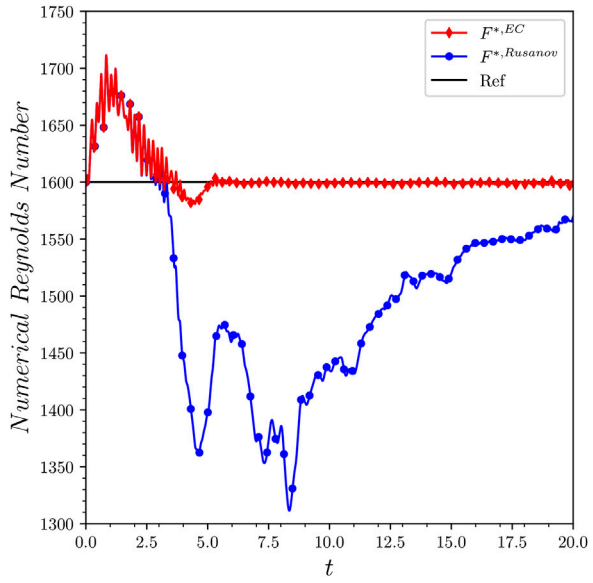


FIGURE 1 | The estimate of the numerical Reynolds number of an under resolved simulation with polynomial degree N = 7 and eight elements (64^3 DOF). The physical Reynolds number of the Taylor-Green vortex setup is Re = 1600. The entropy conservative (EC) scheme retains the physical Reynolds number remarkably well and is virtually dissipation free. The entropy stable (Rusanov) variant clearly introduces stabilizing dissipation as soon as the scales can no longer be resolved starting at about t = 3.

modeled with the ideal MHD equations. For such magnetohydrodynamic flows, the entropy stable DGSEM solver in FLUXO uses a hyperbolic divergence cleaning mechanism to enforce the divergence-free constraint on the magnetic field variables $\vec{B}$ [125]. Additionally, the solver must be augmented with a shock capturing technique in order to handle strong discontinuities [163].

As Io orbits Jupiter, plasma from the torus streams in and forms a local ionosphere that induces polarisation charges and modifies the electric field, thus changing the local Lorentz force and damping electron and ion flow close to Io. The plasma flow is strongly reduced inside Io's (very weak) atmosphere, which can be modeled by incorporating a neutral collision source term to the ideal MHD system, Saur et al. [162, 164],

$$\mathbf{s}_{\mathrm{collision}} = \left[0\,,\ -\varpi\rho\vec{v}\,,\ -\frac{1}{2}\varpi\rho\,\|\,\vec{v}\,\|^2\,,\ \vec{0}\right]^T, \tag{96}$$

with the collision frequency

$$\varpi = \begin{cases} \varpi_{\mathrm{in}} & ,\vec{x}\in \mathfrak{U} \\ \varpi_{\mathrm{in}}\exp\left(\dfrac{r_{\mathfrak{U}}-r}{d}\right) & ,\vec{x}\in \mathfrak{T} \\ 0 & ,\vec{x}\notin \mathfrak{U}\cup\mathfrak{T} \end{cases}, \tag{101}$$

where $\varpi_{\mathrm{in}} > 0$ is constant. The inner atmosphere of Io is represented as a neutral gas cloud $\mathfrak{U}$. In order to model the ionosphere, we also introduce a smooth transition area $\mathfrak{T}$ by an exponential blending dependent on the radii $r_{\mathfrak{U}}, r$ and the dilatation factor d. In this region the neutral atmosphere thins

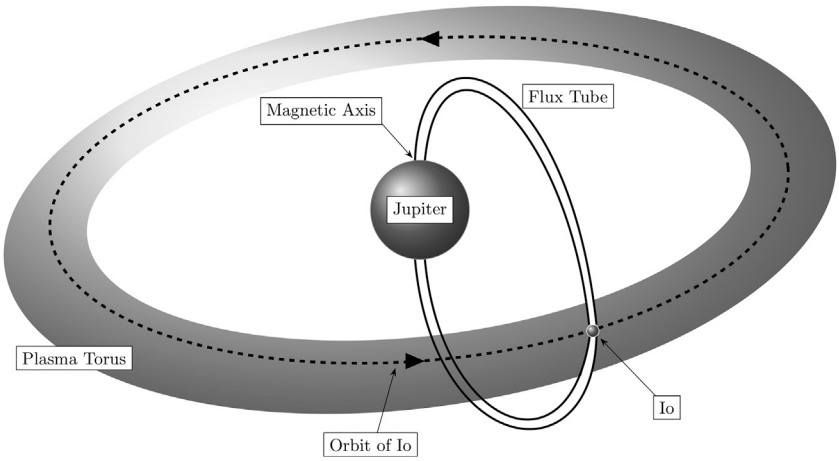


FIGURE 2 | Io's electrodynamic interaction is unique due to its fast rotation and the influence of the strong magnetic field of Jupiter. Plasma interactions with Io's atmosphere lead to mass loss in the form of ions and neutrons. These neutrons then ionize through radiative effects and accumulate around Io forming a plasma torus. Consequently, this flow of magnetized plasma past the obstacle Io, combined with atmospheric interactions, are the engine behind Io's plasma interactions with Jupiter's magnetosphere.

causing ionospheric conductivities to shrink such that they are no longer able to maintain the ionospheric current perpendicular to the magnetic field. Eventually, electric current is continued along the magnetic field lines out of Io's ionosphere, where it is finally fed into Io's Alfvén wings as illustrated in **Figure 3**.

In the dimensionless computational domain we scale the radius of the atmosphere to take a spherical shape with radius one and locate the center of the sphere at the origin

$$\mathfrak{U} = \left\{ \vec{x} \in \Omega \;\middle|\; \| \vec{x} \| \leq r_{\mathfrak{U}} = 1 \right\}. \tag{102}$$

The ionospheric processes uses the exponential blending of the collision frequency above with a dilatation factor $d = 150/1820$. The initial conditions for the flow are taken as

$$\rho = 1, \quad \vec{v} = (1,0,0)^T, \quad p = 0.148, \quad \vec{B} = (0,0,-3.41)^T, \tag{103}$$

that will evolve to a final time $T = 5$. The gas constant is taken to be $\gamma = 5/3$. The boundary states at the left, front and back boundary faces are constant to this reference state, whereas we define outflow boundary conditions at the right, top and bottom of the domain.

In anticipation to capture the relevant physical interactions at the sphere as well as the development of the Alfvén wings best, we exploit the geometric flexibility of the entropy stable DGSEM solver and divide the computational domain into an unstructured, curvilinear mesh presented in **Figure 4**. Within each element we use a polynomial order of $N = 3$.

In **Figure 5** we show a 2D-slice of the B_1 and v_1 components at $y = 0$ of the entropy stable approximation at the final time which presents the numerically generated Alfvén wings from the entropy stable DGSEM. It also demonstrates the expected positive correlation of the B_1 variable with the velocity variable v_1 in the northern Alfvén wing and the negative correlation in the southern wing. Moreover, we consider profile slices in these B_1 and v_1 along the line $z = 5$ and compare the results to a solution computed by the open source software ZEUS (www.astro.princeton.edu/jstone/zeus.html) in **Figure 6**. ZEUS is a FV solver written in spherical coordinates that uses explicit time integration. For the presented comparison, ZEUS used 10 million grid cells (DOF) in total whereas the entropy stable DGSEM solver used approximately 14,336 elements with N=3 polynomials and a total of about 1 million DOF. The reduction of DOFs also translated in a nearly 10 fold reduction of overall CPU time when both codes were run in parallel using MPI on 100 cores. This increased efficiency of the high order DGSEM-LGL, the increased robustness due to discrete entropy stability, and the geometrical flexibility are several advantages of this novel DG framework.

5 WHERE TO GO NEXT?

The response of the DG community to the split form DGSEM with LGL quadrature on tensor-product hexahedra has been astounding. However, naturally, there are still many limitations of this method. Some that have been recently addressed and many others still open. So far, we discussed semi-discrete DGSEM-LGL variants with tensor product expansions on possible curvilinear unstructured hexahedral meshes. Direct extensions of this variant include non-conforming meshes [166, 167], moving meshes [168–170], different related versions such as e.g., the line DG method [171], and a fully discrete space-time approach without the assumption on time continuity [172–175]. An exciting recent development are explicit modified Runge-Kutta methods that retain the semi-discrete entropy stability estimates [176].

A downside of LGL is that in comparison to the Legendre-Gauss (LG) points, the accuracy in dispersion and dissipation is

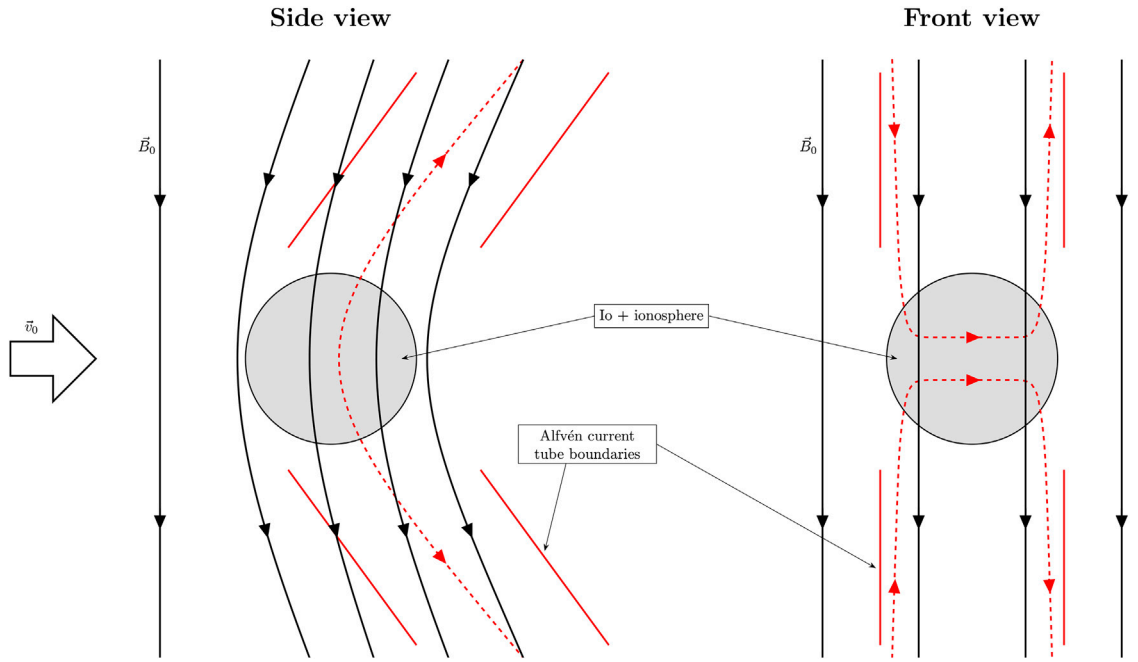


FIGURE 3 | Standing Alfvén current tube magneto-spheric disturbances, termed Alfvén wings, that spread and propagate away along field lines in both directions. The development of the Alfvén wings are observable in the regions with decreasing plasma bulk velocity v_1 and perturbed magnetic field B_1 north and south of Io in the xz plane. It is known that the Alfvén current tubes are bent back by a constant angle with respect to the unperturbed background magnetic field. Further, the perturbation of the magnetic field is positively correlated with the perturbation of the velocity field in the northern Alfvén wing and negatively correlated in the southern wing [133].

lower, see e.g., [56]. Unfortunately, LG points do not include the boundary nodes, hence they do not directly satisfy the classic SBP property and the presented developments cannot be directly applied to this case. However, there are several developments where the framework was extended to construct entropy stable variants with LG nodes. One is based on a staggered grid approach by Parsani et al. [176]. The authors used the standard LG nodes to span the solution, however to compute the discrete derivative operator, they interpolate onto a higher order staggered LGL grid where they can use the SBP property. Then they ensure that the back-projection is still entropy stable to retain the stability estimate on top of the accuracy of the LG nodes. Another approach is presented by e.g., Ortleb [177] where the author directly constructed a scheme that preserves the kinetic energy with LG nodes. Although SBP could not directly be applied, the difference lies only in the boundary operator $\mathcal{B}$. For classic SBP $\mathcal{B}$ is diagonal, whereas in case of LG nodes $\mathcal{B}$ has some columns filled. Ortleb fixed this by considering special correction terms at the boundary, while using similar ideas as in the LGL case for the volume terms. In his PhD thesis, Fernández [178] extended the classic SBP property to general node sets that do or do not include the boundary nodes, where all grid nodes lie inside or even outside of the considered domain.

A generalization onto multi-dimensional domains is given in e.g., [179], termed multidimensional SBP operators. This is an interesting development, as it resolves another limitation of the classic DGSEM-LGL. In some applications, it is favourable to have more flexibility when generating meshes for geometries with complex shapes. In this case, meshes with simplex element types such as triangles and tetrahedra or even hybrid element types such as prisms and pyramids are desired. We will mention some recent developments and extensions here, but stress that this list is not complete and there are many more. Returning to the multidimensional SBP framework, this approach's strength is that it can be used to construct stable methods on simplex meshes e.g., by Chan [115, 180], Chen and Shu [132], Hicken et al. [181], and Crean et al. [182]. Another interesting approach to generate DG scheme on general meshes is presented by Chan [115]. He shows that a special projection directly with the entropy variables collocated at the grid nodes can give a SBP type property that can be used to construct stable discretisations.

As stated the response of the DG community has been astonishing with an explosion of developments, extensions, and new insights. However, there are still many unresolved issues that need to be researched in the future to further evolve high order DG methods into a viable tool for computational physics. The most important problem is still robustness. Although entropy stability significantly improves the stability of the scheme in many applications, there are still situations where no significant gain in robustness of the DG scheme can be observed [183, 184]. A possible reason could be, for instance, the incorrect choice of mathematical entropy function. While there is typically only one physical (thermodynamic) entropy, there are many mathematical ones that often lead to a stability estimate in corresponding norms of the solution. So what are the important entropy quantities to consider? What about e.g., kinetic energy, cross helicity and

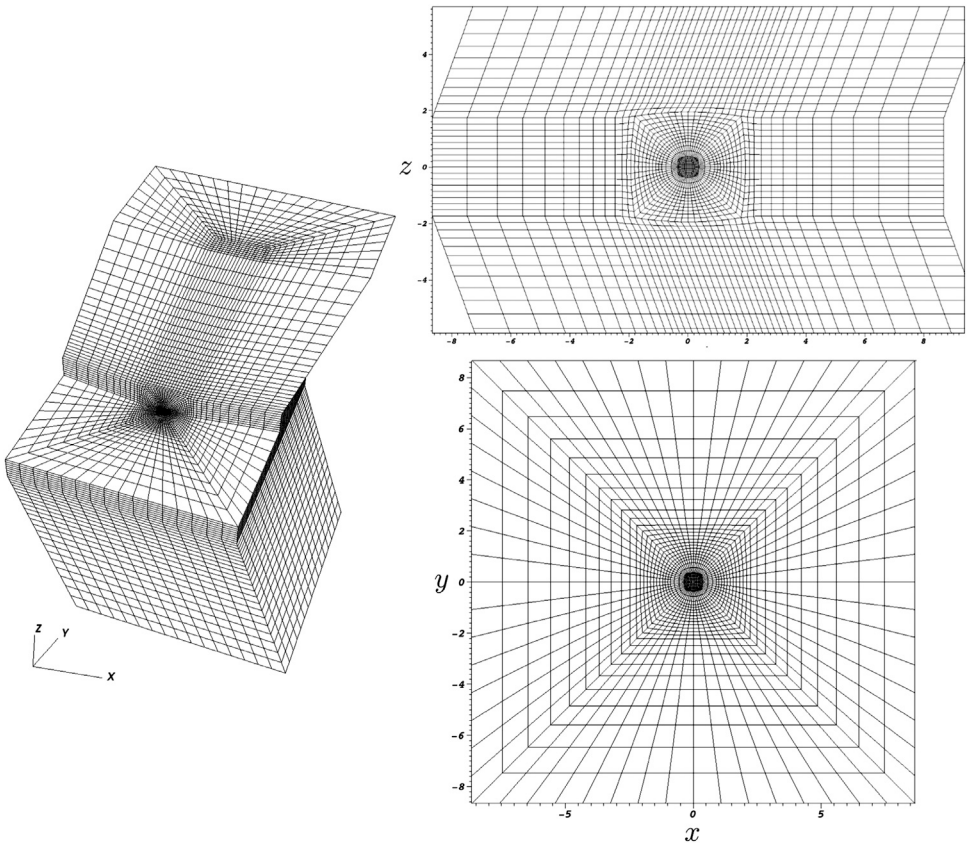


FIGURE 4 | Computational mesh for the Alvén wing test problem built from 14,336 curved elements. At the origin the elements are curved to capture the spherical shape of the neutral gas cloud [11]. Due to the geometric flexibility of the DGSEM the mesh in the northern and southern regions are titled in order to capture the Alvén wing structures more accurately.

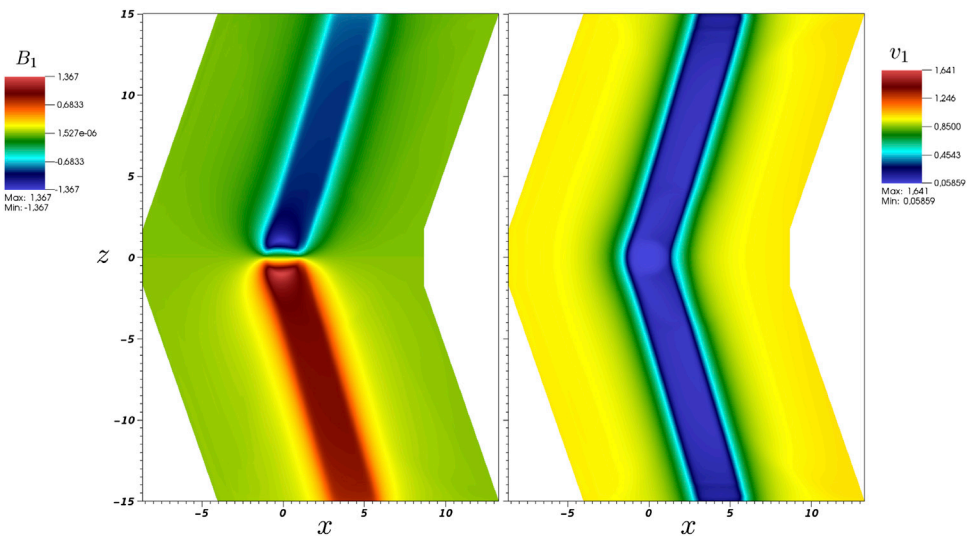


FIGURE 5 | Alfvén wings numerically computed with the entropy stable DGSEM for the plasma interaction of a spherical gas cloud. The snapshot is a slice in the *xz* plane at $y = 0$ at the final time $T = 5$. The polynomial order in each spatial direction was $N = 3$ in each spatial direction. As expected, the Alfvén wings evolve from the northern and southern poles of the neutral gas cloud and are bent back by a constant angle with respect to the background magnetic field. This bending was taken into account in the construction of the curvilinear mesh.

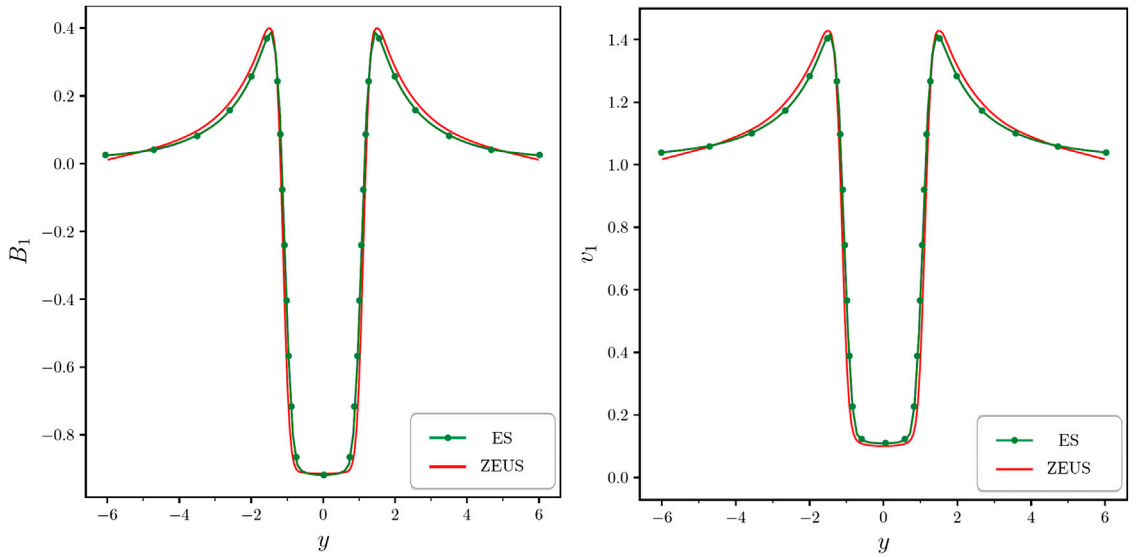


FIGURE 6 | One dimensional visualization of the Alfvén wing solution along the line $z = 5$ from the xz plane slice at $y = 0$. A comparison is performed between the entropy DG solver with $N = 3$ in each spatial direction and the first order finite volume solver ZEUS. The entropy stable DG approximation uses 90% fewer DOF compared to the 10 million DOF used for the ZEUS computation. Qualitatively, the solutions are very similar.

related quantities? For example, in turbulence, the kinetic energy and proper prediction of its behavior seems to play an important role [144, 161, 185, 186]. Besides getting discrete entropy stability estimates, it is possible to use the split form DGSEM-LGL approach from **Section 4.1** to construct DG schemes that discretely preserve the kinetic energy and are discretely compatible for the inner and the total energy, e.g., [116]. It could be shown in Refs. 160 and 161 that such kinetic energy preserving DG schemes behave favourably for the simulation of compressible turbulence in comparison to the standard DG, especially in combination with subgrid turbulence models. However, just as in the development of entropy conservative (and in turn entropy stable) numerical flux functions there is nonuniqeuness. There exist many possible solutions to create a numerical flux that is e.g., kinetic energy preserving or entropy conservative or both e.g., [81, 83, 144, 185, 187]. Moreover, the discrete behavior of the kinetic energy as it evolves in time for under-resolved turbulent flows is quite different even between fluxes that are all provably kinetic energy preserving on paper [116, 187, 188]. So, an important question for the future is thus: What are the important quantities not only from a mathematical, but also from a physical point of view?

A fundamental issue in this context are problem setups that involve physical discontinuities e.g., shock waves in the compressible Euler or ideal MHD equations. Discontinuities trigger another instability, inherent in high order methods: the Gibbs phenomenon i.e., numerical oscillations. These oscillations can be devastating as under- or overshoots can cause non-physical state solutions e.g., negative density or pressure. Hence, positivity is a necessary criterion for all numerical methods when simulating such problems. However, up to this date there is still not enough research into the topic of entropy stability and positivity e.g., [40]. It is worth pointing out, that mathematically, the entropy function is only well-defined for positive solutions and, hence, is strongly connected to positivity. Generalizing this discussion, it is evident that entropy stability is "not enough" as a property for the numerical method. We need more properties, such as e.g., positivity. However, this is also where we reach uncharted research territory as even for many continuous problems e.g., the compressible Navier-Stokes equations it is up to this point unclear to show positivity even for the model itself.

The overview in this work focused on the volume contributions and the underlying tools (physical and mathematical) which led to the entropy stable DG method. However, the contributions at the physical boundaries have been ignored. Properly posing the boundary conditions to be entropy stable for a given model, like the compressible Euler or Navier-Stokes equations, has been considered [189–195], but this is remains an active area of research particularly because the treatment and behavior of the solution (whether on the continuous or discrete level) is directly related to the validity of a mathematical PDE model and directly tied into issues of well-posedness.

Concluding, we are currently at an exciting development stage with high order DG methods, where we can mimic important continuous stability estimates by careful construction of discrete operators. However, practical simulations show that these are not enough for the most complex problems that we desire to simulate. Plus, our numerical schemes and their properties are very close to the current analytical knowledge we have about the physical models. It is very hard to progress with the numerical developments further than what is analytically known: Which properties are important? How do you show positivity? Or a more general question: How does one prove physicality of the solutions for a given PDE model? It seems that the answers can only be given in close collaboration of researchers from physics and mathematics.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

GG was supported by the European Research Council (ERC) under the European Union's Eights Framework Program Horizon 2020 with the research project Extreme, ERC Grant Agreement No. 714487. Support for open access publication was provided by Linköping University.

ACKNOWLEDGMENTS

This work was partially performed on the Cologne High Efficiency Operating Platform for Sciences (CHEOPS) at the Regionales Rechenzentrum Köln (RRZK) at the University of Cologne. The authors thank Marvin Bohm for his contributions to the space physics discussion.

REFERENCES

1. Reed WH, Hill TR. Triangular mesh methods for the neutron transport equation (1973) Technical Report LA-UR-73-479. Los Alamos National Laboratory.
2. Nitsche JA. Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind. *Abh Math Sem Univ Hamburg* (1971) 36:9–15. doi:10.1007/BF02995904
3. Cockburn B, Hou S, Shu C-W. The Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. IV: the multidimensional case. *Math Comput* (1990) 54:545–81. doi:10.2307/2008501
4. Cockburn B, Shu CW. The Runge-Kutta local projection -discontinuous Galerkin method for scalar conservation laws. *Rairo-Matehmatical Model Num Anal-Modelisation Mathematique et Anale Num* (1991) 25:337–61. doi:10.1051/m2an/1991250303371
5. Cockburn B, Shu C-W. The Runge-Kutta discontinuous Galerkin method for conservation laws V: multidimensional systems. *J Comput Phys* (1998) 141: 199–224. doi:10.1006/jcph.1998.5892
6. Cockburn B, Shu CW. Runge-Kutta discontinuous Galerkin methods for convection-dominated problems. *J Sci Comput* (2001) 16:173–261. doi:10.1023/A:1012873910884
7. Bassi F, Rebay S. A high order accurate discontinuous finite element method for the numerical solution of the compressible Navier-stokes equations. *J Comput Phys* (1997) 131:267–79. doi:10.1006/jcph.1996.5572Get
8. Arnold DN, Brezzi F, Cockburn B, Marini D. Discontinuous Galerkin methods for elliptic problems In *Discontinuous Galerkin methods*. Berlin, Heidelberg: Springer (3000). p. 89–101.
9. Arnold DN, Brezzi F, Cockburn B, Marini L. Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM J Numer Anal* (2002) 39:1749–79. doi:10.1137/S0036142901384162
10. Black K. A conservative spectral element method for the approximation of compressible fluid flow. *Kybernetika* (1999) 35:133–46. doi:10.1006/jcph.1996.5572
11. Black K. Spectral element approximation of convection-diffusion type problems. *Appl Numer Math* (2000) 33:373–9. doi:10.1016/S0168-9274(99)00104-X
12. Rasetarinera P, Hussaini M. An efficient implicit discontinuous spectral Galerkin method. *J Comput Phys* (2001) 172:718–38. doi:10.1006/jcph.2001.6853
13. Deng S. Numerical simulation of optical coupling and light propagation in coupled optical resonators with size disorder. *Appl Numer Math* (2007) 57: 475–85. doi:10.1016/j.apnum.2006.07.001
14. Deng S, Cai W, Astratov V. Numerical study of light propagation via whispering gallery modes in microcylinder coupled resonator optical waveguides. *Optic Express* (2004) 12:6468–80. doi:10.1364/opex.12.006468
15. Kopriva DA, Woodruff SL, Hussaini MY. Discontinuous spectral element approximation of Maxwell's Equations In: B Cockburn, G Karniadakis, C-W Shu, editors. Proceedings of the international symposium on discontinuous Galerkin methods. New York: Springer-Verlag (2000) p. 355–61.
16. Kopriva DA, Woodruff SL, Hussaini MY. Computation of electromagnetic scattering with a non-conforming discontinuous spectral element method. *Int J Numer Methods Eng* (2002) 53:105–22. doi:10.1002/nme.394
17. Chan J, Hewett RJ, Warburton T. Weight-adjusted discontinuous Galerkin methods: wave propogation in heterogeneous media. *SIAM J Sci Comput* (2017) 39:A2935–A2961. doi:10.1002/nme.5720
18. Rasetarinera P, Kopriva D, Hussaini M. Discontinuous spectral element solution of acoustic radiation from thin airfoils. *AIAA J* (2001) 39:2070–5. doi:10.2514/3.14970
19. Stanescu D, Farassat F, Hussaini M. Aircraft Engine noise scattering – parallel discontinuous Galerkin spectral element method. *AIAA* (2002a) doi:10.2514/6.2002-800
20. Stanescu D, Xu J, Farassat F, Hussaini M. Computation of engine noise propagation and scattering off an aircraft. *Aeroacoustics* (2002b) 1:403–20. doi:10.1260/147547202765275989
21. Wilcox LC, Stadler G, Burstedde C, Ghattas O. A high-order discontinuous Galerkin method for wave propagation through coupled elastic-acoustic media. *J Comput Phys* (2010) 229:9373–96. doi:10.1016/j.jcp.2010.09.008
22. Bonev B, Hesthaven JS, Giraldo FX, Kopera MA. Discontinuous Galerkin scheme for the spherical shallow water equations with applications to tsunami modeling and prediction. *J Comput Phys* (2018) 362:425–48. doi:10.1016/j.jcp.2018.02.008
23. Giraldo F, Hesthaven J, Warburton T. Nodal high-order discontinuous Galerkin methods for the spherical shallow water equations. *J Comput Phys* (2002) 181:499–525. doi:10.1006/jcph.2002.7139
24. Giraldo F, Restelli M. A study of spectral element and discontinuous Galerkin methods for the Navier–Stokes equations in nonhydrostatic mesoscale atmospheric modeling: equation sets and test cases. *J Comput Phys* (2008) 227:3849–77. doi:10.1016/j.jcp.2007.12.009
25. Restelli M, Giraldo F. A conservative discontinuous Galerkin semi-implicit formulation for the Navier-Stokes equations in nonhydrostatic mesoscale modeling. *SIAM J Sci Comput* (2009) 31:2231–57. doi:10.1137/070708470
26. Fagherazzi S, Furbish D, Rasetarinera P, Hussaini MY. Application of the discontinuous spectral Galerkin method to groundwater flow. *Adv Water Res* (2004a) 27:129–40. doi:10.1016/j.advwatres.2003.11.001
27. Fagherazzi S, Rasetarinera P, Hussaini MY, Furbish DJ. Numerical solution of the dam-break problem with a discontinuous Galerkin method. *J Hydraul Eng* (2004b) 130:532–9. doi:10.1061/(ASCE)0733
28. Cockburn B, Karniadakis G, Shu C-W. The development of discontinuous Galerkin methods. In: B Cockburn, G Karniadakis, C-W Shu, editors. Proceedings of the inter)national symposium on discontinuous Galerkin methods. New York: Springer-Verlag (2000). p. 3–50.
29. Hesthaven J, Warburton T. *Nodal discontinuous Galerkin methods*: algorithms, analysis, and applications. Verlag New York: Springer (2008)
30. Karniadakis GE, Sherwin SJ. *Spectral/hp element methods for computational fluid dynamics*. Oxford:Oxford University Press (2005).
31. Kopriva DA. *Implementing spectral methods for partial differential equations*. Scientific Computation (Netherlands:Springer) (2009).
32. Gassner GJ, Winters AR, Hindenlang FJ, Kopriva DA. The BR1 scheme is stable for the compressible Navier-Stokes equations. *J Sci Comput* (2018) 77: 154–200.
33. Toro EF. *Riemann solvers and numerical methods for fluid dynamics*. Springer-Verlag (1999).
34. Kopriva DA, Gassner G. On the quadrature and weak form choices in collocation type discontinuous Galerkin spectral element methods. *J Sci Comput* (2010) 44:136–55.

35. Kuzmin D. Entropy stabilization and property-preserving limiters for discontinuous Galerkin discretizations of scalar hyperbolic problems. *J Numer Math* (2020) 1. doi:10.1515/jnma-2020-0056
36. Guo W, Nair RD, Zhong X. An efficient WENO limiter for discontinuous Galerkin transport scheme on the cubed sphere. *Int J Numer Methods Fluid* (2016) 81:3–21.
37. Zhu J, Qiu J. Hermite WENO schemes and their application as limiters for Runge-Kutta discontinuous Galerkin method, III: unstructured meshes. *J Sci Comput* (2009) 39:293–321.
38. Bohm M, Schermeng S, Winters AR, Gassner GJ, Jacobs GB. Multi-element SIAC filter for shock capturing applied to high-order discontinuous Galerkin spectral element methods. *J Sci Comput* (2019) 81:820–44.
39. Gassner GJ, Beck AD. On the accuracy of high-order discretizations for underresolved turbulence simulations. *Theor Comput Fluid Dynam* (2013) 27:221–37.
40. Hennemann S, Rueda-Ramírez AM, Hindenlang FJ, Gassner GJ (2020). A provably entropy stable subcell shock capturing approach for high order split form DG for the compressible Euler equations. *J Comput Phys* 109935. doi:10.1016/j.jcp.2020.109935
41. Markert J, Gassner G, Walch S (2020). A sub-element adaptive shock capturing approach for discontinuous Galerkin methods. arXiv:2011.03338.
42. Sonntag M, Munz C-D. Efficient parallelization of a shock capturing for discontinuous Galerkin methods using finite volume sub-cells. *J Sci Comput* (2017) 70:1262–89. doi:10.1007/s10915-016-0287-5
43. Dumbser M, Loubère R. A simple robust and accurate *a posteriori* sub-cell finite volume limiter for the discontinuous Galerkin method on unstructured meshes. *J Comput Phys* (2016) 319:163–99. doi:10.1016/j.jcp.2016.05.002
44. Dumbser M, Zanotti O, Loubère R, Diot S. *A posteriori* subcell limiting of the discontinuous Galerkin finite element method for hyperbolic conservation laws. *J Comput Phys* (2014) 278:47–75. doi:10.1016/j.jcp.2014.08.009
45. Fambri F. Discontinuous Galerkin methods for compressible and incompressible flows on space–time adaptive meshes: toward a novel family of efficient numerical methods for fluid dynamics. *Arch Comput Methods Eng* (2020) 27:199–283. doi:10.1007/S11831-018-09308-6
46. Giri P, Qiu J. A high-order Runge-Kutta discontinuous Galerkin method with a subcell limiter on adaptive unstructured grids for two-dimensional compressible inviscid flows. *Int J Numer Methods Fluid* (2019) 91:367–94. doi:10.1002/fld.4757
47. Zanotti O, Fambri F, Dumbser M, Hidalgo A. Space–time adaptive ADER discontinuous Galerkin finite element schemes with *a posteriori* sub-cell finite volume limiting. *Comput. Fluids* (2015) 118:204–24. doi:10.1016/j.compfluid.2015.06.020
48. Ching EJ, Lv Y, Gnoffo P, Barnhardt M, Ihme M. Shock capturing for discontinuous Galerkin methods with application to predicting heat transfer in hypersonic flows. *J Comput Phys* (2019) 376:54–75. doi:10.1016/j.jcp.2018.09.016
49. Persson P-O, Peraire J. Sub-cell shock capturing for discontinuous Galerkin methods. In 44th AIAA aerospace sciences meeting and exhibit (2006). p. 112.
50. Beskok A, Warburton TC. An unstructured finite-element scheme for fluid flow and heat transfer in moving domains. *J Comput Phys* (2001) 174:492–509. doi:10.1006/jcph.2001.6885
51. Carpenter M, Kennedy C, Bijl H, Viken S, Vatsa V. Fourth-order Runge-Kutta schemes for fluid mechanics applications. *J Sci Comput* (2005) 25:157–94. doi:10.1007/s10915-004-4637-3
52. Constantinescu E, Sandu A. Multirate explicit adams methods for time integration of conservation laws. *J Sci Comput* (2009) 38:229–49. doi:10.1007/s10915-008-9235-3
53. Gassner G, Haas M. An explicit high order accurate predictor-corrector time integration method with consistent local time-stepping for discontinuous Galerkin schemes. *AIP Conf Proceed* (2009) 1168:1188–91. doi:10.1063/1.3241276
54. Kopera MA, Giraldo FX. Analysis of adaptive mesh refinement for IMEX discontinuous Galerkin solutions of the compressible Euler equations with application to atmospheric simulations. *J Comput Phys* (2014) 275:92–117. doi:10.1016/j.jcp.2014.06.026
55. Gassner G, Kopriva DA. A comparison of the dispersion and dissipation errors of Gauss and Gauss-Lobatto discontinuous Galerkin spectral element methods. *SIAM J Sci Comput* (2010) 33:2560–79. doi:10.1137/100807211
56. Stanescu D, Kopriva D, Hussaini M. Dispersion analysis for discontinuous spectral element methods. *J Sci Comput* (2001) 15:149–71. doi:10.1023/A:1007629609576
57. Altmann C, Beck AD, Hindenlang F, Staudenmaier M, Gassner GJ, Munz C-D. An efficient high performance parallelization of a discontinuous Galerkin spectral element method. In: R Keller, D Kramer, J-P Weiss, editors *Facing the multicore-challenge IIIof lecture notes in computer science*. Springer Berlin Heidelberg (2013), Vol. 7686, p. 37–47.
58. Hindenlang F, Gassner GJ, Altmann C, Beck A, Staudenmaier M, Munz C-D. Explicit discontinuous Galerkin methods for unsteady problems. *Comput Fluids* (2012) 61:86–93. doi:10.1016/j.compfluid.2012.03.006
59. Dumbser M, Käser M, Toro EF. An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes–V. local time stepping and *p*-adaptivity. *Geophys J Int* (2007) 171:695–717. doi:10.1111/j.1365-246X.2007.03427.x
60. Frank HM, Munz C-D. Direct aeroacoustic simulation of acoustic feedback phenomena on a side-view mirror. *J Sound Vib* (2016) 371:132–49. doi:10.1016/j.jsv.2016.02.014
61. Persson P-O. A sparse and high-order accurate line-based discontinuous Galerkin method for unstructured meshes. *J Comput Phys* (2013) 233:414–29. doi:10.1016/j.jcp.2012.09.008
62. Warburton T. Application of the discontinuous Galerkin method to Maxwell's equations using unstructured polymorphic -finite elements. In: B Cockburn, G Karniadakis, C-W Shu, editors. Proceedings of the international symposium on discontinuous Galerkin methods. New York: Springer-Verlag (2000). p. 451–8.
63. Wintermeyer N, Winters AR, Gassner GJ, Kopriva DA. An entropy stable nodal discontinuous Galerkin method for the two dimensional shallow water equations on unstructured curvilinear meshes with discontinuous bathymetry. *J Comput Phys* (2017) 340:200–42. doi:10.1016/j.jcp.2017.03.036
64. Kopriva DA, Gassner GJ. An energy stable discontinuous Galerkin spectral element discretization for variable coefficient advection problems. *SIAM J Sci Comput* (2014) 34:A2076–A2099. doi:10.1137/130928650
65. Jiang G, Shu C-W. On a cell entropy inequality for discontinuous Galerkin methods. *Math Comput* (1994) 62:531–8. doi:10.2307/2153521
66. Kopriva DA. Stability of overintegration methods for nodal discontinuous Galerkin spectral element methods. *J Sci Comput* (2017b) 76:426–42. doi:10.1007/s10915-017-0626-1
67. Moura R, Sherwin S, Peiro J. On DG-based iLES approaches at very high Reynolds numbers. Research Gate: Report (2015).
68. Merriam ML. Smoothing and the second law. *Comput Methods Appl Mech Eng* (1987) 64:177–93. doi:10.1016/0045-7825(87)90039-9
69. Chiodaroli E. A counterexample to well-posedness of entropy solutions to the compressible Euler system. *J Hyperbolic Differ Equ* (2014) 11:493–519. doi:10.1142/S0219891614500143
70. Klingenberg C, Markfelder S. Non-uniqueness of entropy-conserving solutions to the ideal compressible MHD equations. In: A Bressan, M Lewicka, D Wang, Y Zheng, editors. *Hyperbolic problems: theory, numerics, applications*. AIMS on Applied Mathematics (2020), Vol. 10, p. 491–8.
71. LeFloch PG, Novotny J. Hyperbolic systems of conservation laws: the theory of classical and nonclassical shock waves. *Appl Mech Rev* (2003) 56:B53–4. doi:10.1007/978-3-0348-8150-0
72. Terracina A. Non-uniqueness results for entropy two-phase solutions of forward–backward parabolic problems with unstable phase. *J Math Anal Appl* (2014) 413:963–75. doi:10.1016/j.jmaa.2013.12.045
73. Evans LC. *Partial differential equations*. Providence, Rhode Island:American Mathematical Society (2012).
74. LeVeque RJ. *Finite volume methods for hyperbolic problems*. Cambridge: Cambridge University Press (2002).
75. Lax PD. Weak solutions of nonlinear hyperbolic conservation equations and their numerical computation. *Commun Pure Appl Math* (1954) 7:159–93. 10.1002/cpa.3160070112
76. Lax PD. Hyperbolic difference equations: a review of the Courant-Friedrichs-Lewy paper in the light of recent developments. *IBM J Res Develop* (1967) 11:235–8. doi:10.1147/rd.112.0235
77. Tadmor E. Entropy functions for symmetric systems of conservation laws. *J Math Anal Appl* (1987) 122:355–9. doi:10.1016/0022-247X(87)90265-4

78. Tadmor E. Entropy stability theory for difference approximations of nonlinear conservation laws and related time-dependent problems. *Acta Numer* (2003) 12:451–512. doi:10.1017/S0962492902000156
79. Ranocha H. Mimetic properties of difference operators: product and chain rules as for functions of bounded variation and entropy stability of second derivatives. *BIT Num Math* (2019) 59:547–63. doi:10.1007/s10543-018-0736-7
80. Harten A. On the symmetric form of systems of conservation laws with entropy. *J Comput Phys* (1983) 49:151–64. doi:10.1016/0021-9991(83)90118-3
81. Chandrashekar P. Kinetic energy preserving and entropy stable finite volume schemes for compressible Euler and Navier-Stokes equations. *Commun Comput Phys* (2013) 14:1252–86. doi:10.4208/cicp.170712.010313a
82. Dutt P. Stable boundary conditions and difference schemes for Navier-Stokes equations. *SIAM J Numer Anal* (1988) 25:245–67. doi:10.1137/0725018
83. Ismail F, Roe PL. Affordable, entropy-consistent Euler flux functions II: entropy production at shocks. *J Comput Phys* (2009) 228:5410–36. doi:10.1016/j.jcp.2009.04.021
84. Mock MS. Systems of conservation laws of mixed type. *J Differ Equ* (1980) 37:70–88. doi:10.1016/0022-0396(80)90089-3
85. Tadmor E. Skew-selfadjoint form for systems of conservation laws. *J Math Anal Appl* (1984) 103:428–42. doi:10.1016/0022-247X(84)90139-2
86. Merriam ML. An entropy-based approach to nonlinear stability. *NASA Tech Memo* (1989) 101086:1–154.
87. Derigs D, Gassner GJ, Walch S, Winters AR. Entropy stable finite volume approximations for ideal magnetohydrodynamics. *Jahresber Dtsch Math Ver* (2018) 120:153–219. doi:10.1365/s13291-018-0178-9
88. Fisher T. *High-order stable multi-domain finite difference method for compressible flows*. [PhD thesis]. West Lafayette, Indiana:Purdue University (2012).
89. Evans LC. *Entropy and partial differential equations*. Lecture Notes at Berkeley, California:University of California, Berkeley (2004).
90. Svärd M. Entropy solutions of the compressible Euler equations. *BIT Num Math* (2016) 56:1479–96. doi:10.1007/s10543-016-0611-3
91. Hughes TJ, Franca L, Mallet M. A new finite element formulation for computational fluid dynamics: I. Symmetric forms of the compressible Euler and Navier-Stokes equations and the second law of thermodynamics. *Comput Methods Appl Mech Eng* (1986) 54:223–34. doi:10.1016/0045-7825(86)90127-1
92. Hiltebrand A, Mishra S. Entropy stable shock capturing space–time discontinuous Galerkin schemes for systems of conservation laws. *Numer Math* (2013) 126:103–51. doi:10.1007/s00211-013-0558-0
93. Hiltebrand A, Mishra S. Entropy stability and well-balancedness of space-time DG for the shallow water equations with bottom topography. *Netw Heterogeneous Media* (2016) 11:145–62. doi:10.3934/nhm.2016.11.145
94. Hiltebrand A, Mishra S, Parés C. Entropy-stable space–time DG schemes for non-conservative hyperbolic systems. *ESAIM Math Model Numer Anal* (2018) 52:995–1022. doi:10.1051/m2an/2017056
95. Kirby RM, Karniadakis G. De-aliasing on non-uniform grids: algorithms and applications. *J Comput Phys* (2003) 191:249–64. doi:10.1016/S0021-9991(03)00314-0
96. Mengaldo G, Grazia DD, Moxey D, Vincent PE, Sherwin SJ. Dealiasing techniques for high-order spectral element methods on regular and irregular grids. *J Comput Phys* (2015) 299:56–81. doi:10.1016/j.jcp.2015.06.032
97. Beck AD, Bolemann T, Flad D, Frank H, Gassner GJ, Hindenlang F, et al. High-order discontinuous Galerkin spectral element methods for transitional and turbulent flow simulations. *Int J Numer Methods Fluid* (2014) 76:522–48. doi:10.1002/fld.3943
98. Persson P-O, Bonet J, Peraire J. Discontinuous Galerkin solution of the Navier-Stokes equations on deformable domains. *Comput Methods Appl Mech Eng* (2009) 198:1585–95. doi:10.1016/j.cma.2009.01.012
99. LeFloch P, Rohde C. High-order schemes, entropy inequalities, and nonclassical shocks. *SIAM J Numer Anal* (2000) 37:2023–60. doi:10.1137/S0036142998345256
100. LeFloch PG, Mercier J-M, Rohde C. Fully discrete, entropy conservative schemes of arbitrary order. *SIAM J Numer Anal* (2002) 40:1968–92. doi:10.1137/S003614290240069X
101. Fisher T, Carpenter M. High-order entropy stable finite difference schemes for nonlinear conservation laws: finite domains. *J Comput Phys* (2013) 252:518–57. doi:10.1016/j.jcp.2013.06.014
102. Kreiss H-O, Oliger J. Methods for the approximate solution of time-dependent problems (1973). GARP Report No.:10. Geneva: World Meteorological Organization.
103. Kreiss H-O, Olliger J. Comparison of accurate methods for the integration of hyperbolic equations. *Tellus* (1972) 24:199–215. doi:10.3402/tellusa.v24i3.10634
104. Kreiss H-O, Scherer G. Finite element and finite difference methods for hyperbolic partial differential equations. *Math Aspect Finite Elem Part Diff Equat (Elsevier)* (1974) 195–212. doi:10.1016/B978-0-12-208350-1.50012-1
105. Kreiss H-O, Scherer G. On the existence of energy estimates for difference approximations for hyperbolic systems (1977). Tech. Rep. Uppsala: Department of Scientific Computing. Uppsala University.
106. Olsson P. Summation by parts, projections, and stability. I. *Math Comput* (1995a) 64:1035–65. doi:10.2307/2153482
107. Olsson P. Summation by parts, projections, and stability. II. *Math Comput* (1995b) 64:1473–93. doi:10.2307/2153366
108. Strand B. Summation by parts for finite difference approximations for d/d*x*. *J Comput Phys* (1994) 110:47–67. doi:10.1006/jcph.1994.1005
109. Nordström J. Conservative finite difference formulations, variable coefficients, energy estimates and artificial dissipation. *J Sci Comput* (2006) 29:375–404. doi:10.1007/s10915-005-9013-4
110. Svärd M, Nordström J. Review of summation-by-parts schemes for initial-boundary-value problems. *J Comput Phys* (2014) 268:17–38. doi:10.1016/j.jcp.2014.02.031
111. Gassner G. A skew-symmetric discontinuous Galerkin spectral element discretization and its relation to SBP-SAT finite difference methods. *SIAM J Sci Comput* (2013) 35:A1233–A1253. doi:10.1137/120890144
112. Carpenter MH, Gottlieb D. Spectral methods on arbitrary grids. *J Comput Phys* (1996) 129:74–86. doi:10.1006/jcph.1996.0234
113. Kopriva DA. A polynomial spectral calculus for analysis of DG spectral element methods. In Bittencourt ML, Dumont NA, and Hesthaven, JS, editors. Spectral and high order methods for partial differential equations ICOSAHOM 2016. Cham:Springer International Publishing (2017) p. 21–40.
114. Carpenter M, Fisher T, Nielsen E, Frankel S. Entropy stable spectral collocation schemes for the Navier-Stokes equations: discontinuous interfaces. *SIAM J Sci Comput* (2014) 36:B835–B867. doi:10.1137/130932193
115. Chan J. On discretely entropy conservative and entropy stable discontinuous Galerkin methods. *J Comput Phys* (2018) 362:346–74. doi:10.1016/j.jcp.2018.02.033
116. Gassner G, Winters A, Kopriva DA. Split form nodal discontinuous Galerkin schemes with summation-by-parts property for the compressible Euler equations. *J Comput Phys* (2016a) 327:39–66. doi:10.1016/j.jcp.2016.09.013
117. Gassner GJ. A kinetic energy preserving nodal discontinuous Galerkin spectral element method. *Int J Numer Methods Fluid* (2014) 76:28–50. doi:10.1002/fld.3923
118. Gouasmi A, Duraisamy K, Murman SM. Formulation of entropy-stable schemes for the multicomponent compressible Euler equations. *Comput Methods Appl Mech Eng* (2020) 363:112912. doi:10.1016/j.cma.2020.112912
119. Gassner GJ, Winters AR, Kopriva DA. A well balanced and entropy conservative discontinuous Galerkin spectral element method for the shallow water equations. *Appl Math Comput* (2016b) 272(2):291–308. doi:10.1016/j.amc.2015.07.014
120. Wu X, Chan J (2020). Entropy stable discontinuous Galerkin methods for nonlinear conservation laws on networks and multi-dimensional domains. arXiv:2010.09994.
121. Wu X, Chan J, Kubatko E (2020). High-order entropy stable discontinuous Galerkin methods for the shallow water equations: curved triangular meshes and GPU acceleration. arXiv:2005.02516.
122. Parsani M, Carpenter M, Nielsen E. Entropy stable discontinuous interfaces coupling for the three-dimensional compressible Navier-Stokes equations. *J Comput Phys* (2015a) 290:132–8. doi:10.1016/j.jcp.2015.02.042
123. Parsani M, Carpenter M, Nielsen E. Entropy stable wall boundary conditions for the three-dimensional compressible Navier-Stokes equations. *J Comput Phys* (2015b) 292:88–113. doi:10.1016/j.jcp.2015.03.026
124. Renac F. Entropy stable DGSEM for nonlinear hyperbolic systems in nonconservative form with application to two-phase flows. *J Comput Phys* (2019) 382:1–26. doi:10.1016/j.jcp.2018.12.035
125. Bohm M, Winters AR, Gassner GJ, Derigs D, Hindenlang F, Saur J. An entropy stable nodal discontinuous Galerkin method for the resistive MHD equations. Part I: theory and numerical verification. *J Comput Phys* (2018) 422:108076. doi:10.1016/j.jcp.2018.06.027

126. Liu Y, Shu C-W, Zhang M. Entropy stable high order discontinuous Galerkin methods for ideal compressible MHD on structured meshes. *J Comput Phys* (2018) 354:163–78. doi:10.1016/j.jcp.2017.10.043
127. Biswas B, Kumar H (2019). Entropy stable discontinuous Galerkin approximation for the relativistic hydrodynamic equations. arXiv: 1911.07488.
128. Wu K, Shu C-W (2019). Entropy symmetrization and high-order accurate entropy stable numerical schemes for relativistic MHD equations. arXiv: 1907.07467.
129. Manzanero J, Rubio G, Kopriva DA, Ferrer E, Valero E. A free–energy stable nodal discontinuous Galerkin approximation with summation–by–parts property for the Cahn–Hilliard equation. *J Comput Phys* (2020c) 403: 109072. doi:10.1016/j.jcp.2019.109072
130. Manzanero J, Rubio G, Kopriva DA, Ferrer E, Valero E. An entropy–stable discontinuous Galerkin approximation for the incompressible Navier–Stokes equations with variable density and artificial compressibility. *J Comput Phys* (2020a) 408:109241. doi:10.1016/j.jcp.2020.109241
131. Manzanero J, Rubio G, Kopriva DA, Ferrer E, Valero E. Entropy–stable discontinuous Galerkin approximation with summation–by–parts property for the incompressible Navier–Stokes/Cahn–Hilliard system. *J Comput Phys* (2020b) 408:109363. doi:10.1016/j.jcp.2020.109363
132. Chen T, Shu C-W. Entropy stable high order discontinuous Galerkin methods with suitable quadrature rules for hyperbolic conservation laws. *J Comput Phys* (2017) 345:427–61. doi:10.1016/j.jcp.2017.05.025
133. Zang TA. On the rotation and skew-symmetric forms for incompressible flow simulations. *Appl Numer Math* (1991) 7:27–40. doi:10.1016/0168-9274(91)90102-6
134. Ducros F, Laporte F, Soulères T, Guinot V, Moinat P, Caruelle B. High-order fluxes for conservative skew-symmetric-like schemes in structured meshes: application to compressible flows. *J Comput Phys* (2000) 161:114–39. doi:10.1006/jcph.2000.6492
135. Honein AE, Moin P. Higher entropy conservation and numerical stability of compressible turbulence simulations. *J Comput Phys* (2004) 201:531–45. doi:10.1016/j.jcp.2004.06.006
136. Kennedy CA, Gruber A. Reduced aliasing formulations of the convective terms within the Navier–Stokes equations for a compressible fluid. *J Comput Phys* (2008) 227:1676–700. doi:10.1016/j.jcp.2007.09.020
137. Pirozzoli S. Generalized conservative approximations of split convective derivative operators. *J Comput Phys* (2010) 229:7180–90. doi:10.1016/j.jcp.2010.06.006
138. Sandham ND, Li Q, Yee HC. Entropy splitting for high-order numerical simulation of compressible turbulence. *J Comput Phys* (2002) 178:307–22. doi:10.1006/jcph.2002.7022
139. Sjögreen B, Yee HC, Kotov D. Skew-symmetric splitting and stability of high order central schemes. *J Phys Conf* (2017) 837:012019. doi:10.1007/s00193-019-00925-z
140. Blaisdell GA, Spyropoulos ET, Qin JH. The effect of the formulation of nonlinear terms on aliasing errors in spectral methods. *Appl Numer Math* (1996) 21:207–19. doi:10.1016/0168-9274(96)00005-0
141. LeFloch PG, Mohammadian M. Why many theories of shock waves are necessary: kinetic functions, equivalent equations, and fourth-order models. *J Comput Phys* (2008) 227:4162–89. doi:10.1016/j.jcp.2007.12.026
142. Winters AR, Gassner GJ. A comparison of two entropy stable discontinuous Galerkin spectral element approximations for the shallow water equations with non-constant topography. *J Comput Phys* (2015) 301:357–76. doi:10.1016/j.jcp.2015.08.034
143. Coppola G, Capuno F, Pirozzoli S, de Luce L. Numerically stable formulations of convective terms for turbulent compressible flows. *J Comput Phys* (2019) 382:86–104. doi:10.1016/j.jcp.2019.01.007
144. Kuya Y, Totani K, Kawai S. Kinetic energy and entropy preserving schemes for compressible flows by split convective forms. *J Comput Phys* (2018) 375: 823–53. doi:10.1016/j.jcp.2018.08.058
145. Tadmor E. Perfect derivatives, conservative differences and entropy stable computation of hyperbolic conservation laws. *Disc Contin Dyn Syst-A* (2016) 36:4579–98. 10.3934/dcds.2016.36.4579
146. Fjordholm US, Mishra S, Tadmor E. Well-balanced and energy stable schemes for the shallow water equations with discontiuous topography. *J Comput Phys* (2011) 230:5587–609. doi:10.1016/j.jcp.2011.03.042
147. Winters A, Gassner G. Affordable, entropy conserving and entropy stable flux functions for the ideal MHD equations. *J Comput Phys* (2016) 301:72–108. doi:10.1016/j.jcp.2015.09.055
148. Fjordholm US, Mishra S, Tadmor E. Arbitrarily high-order accurate entropy stable essentially nonoscillatory schemes for systems of conservation laws. *SIAM J Numer Anal* (2012) 50:544–73. doi:10.1137/110836961
149. Fjordholm US, Ray D. A sign preserving WENO reconstruction method. *J Sci Comput* (2016) 68:42–63. doi:10.1007/s10915-015-0128-y
150. Ranocha H. Shallow water equations: split-form, entropy stable, well-balanced, and positivity preserving numerical methods. *GEM-International Journal on Geomathematics* (2017) 8:85–133. doi:10.1007/s13137-016-0089-9
151. Ranocha H. Comparison of some entropy conservative numerical fluxes for the Euler equations. *J Sci Comput* (2018) 76:216–42. doi:10.1007/s10915-017-0618-1
152. Barth TJ. Numerical methods for gasdynamic systems on unstructured meshes. In: D Kröner, M Ohlberger, C Rohde, editors *An introduction to recent developments in theory and numerics for conservation laws of lecture notes in computational science and engineering*. Springer Berlin Heidelberg (1999), Vol. 5, p. 195–285.
153. Derigs D, Winters AR, Gassner GJ, Walch S. A novel averaging technique for discrete entropy-stable dissipation operators for ideal MHD. *J Comput Phys* (2017) 330:624–32. doi:10.1016/j.jcp.2016.10.055
154. Fjordholm US. *High-order accurate entropy stable numercial schemes for hyperbolic conservation laws*. [Ph.D. thesis]. ETH Zurich (2013).
155. Ray D, Chandrashekar P. An entropy stable finite volume scheme for the two dimensional Navier–Stokes equations on triangular grids. *Appl Math Comput* (2017) 314:257–86. doi:10.1016/j.amc.2017.07.020
156. Winters AR, Derigs D, Gassner GJ, Walch S. A uniquely defined entropy stable matrix dissipation operator for high Mach number ideal MHD and compressible Euler simulations. *J Comput Phys* (2017) 332:274–89. doi:10.1016/j.jcp.2016.12.006
157. Chan J, Wilcox LC. On discretely entropy stable weight-adjusted discontinuous Galerkin methods: curvilinear meshes. *J Comput Phys* (2019) 378:366–93. doi:10.1016/j.jcp.2018.11.010
158. Rojas D, Boukharfane R, Dalcin L, Fernández DCDR, Ranocha H, Keyes DE, et al. On the robustness and performance of entropy stable collocated discontinuous Galerkin methods. *J Comput Phys* (2020) 109891. doi:10.1016/j.jcp.2020.109891
159. Carpenter M, Kennedy C. Fourth-order -storage Runge-Kutta schemes (1994). Tech. Rep. NASA TM 109111. Hampton, Virginia:NASA Langley Research Center.
160. Winters AR, Moura RC, Mengaldo G, Gassner GJ, Walch S, Peiro J, et al. A comparative study on polynomial dealiasing and split form discontinuous Galerkin schemes for under-resolved turbulence computations. *J Comput Phys* (2018) 372:1–21. doi:10.1016/j.jcp.2018.06.016
161. Flad D, Gassner G. On the use of kinetic energy preservaing DG-scheme for large eddy simulation. *J Comput Phys* (2017) 350:782–95. doi:10.1016/j.jcp.2020.109891
162. Saur J, Neubauer FM, Strobel DF, Summers ME. Three-dimensional plasma simulation of Io's interaction with the Io plasma torus: asymmetric plasma flow. *J Geophys Res: Space Physics* (1999) 104:25105–26. doi:10.1029/1999JA900304
163. Bohm M. *An entropy stable nodal discontinuous Galerkin method for the resistive MHD equations*. [PhD thesis]. Universität zu Köln (2018).
164. Saur J, Neubauer FM, Connerney J, Zarka P, Kivelson MG. Plasma interaction of Io with its plasma torus. In: Bagenal F, Dowling, TE, and McKinnon, WB, editors. Jupiter. The Planet, Satellites and Magnetosphere. vol. 1. Cambridge: Cambridge University Press (2004). p. 537–60.
165. Chan J, Bencomo M, Fernández DCDR (2020). Mortar-based entropy-stable discontinuous Galerkin methods on non-conforming quadrilateral and hexahedral meshes. arXiv:2005.03237.
166. Friedrich L, Winters AR, Fernández DCDR, Gassner GJ, Parsani M, Carpenter MH. An entropy stable non-conforming discontinuous Galerkin method with the summation-by-parts property. *J Sci Comput* (2018) 77:689–725. doi:10.1007/s10915-018-0733-7
167. Kopriva DA, Winters AR, Bohm M, Gassner GJ. A provably stable discontinuous Galerkin spectral element approximation for moving

hexahedral meshes. *Comput Fluids* (2016) 139:148–60. doi:10.1016/j.compfluid.2016.05.023
168. Schnücke G, Krais N, Bolemann T, Gassner GJ. Entropy stable discontinuous Galerkin schemes on moving meshes for hyperbolic conservation laws. *J Sci Comput* (2020) 82:1–42. doi:10.1007/s10915-020-01171-7
169. Yamaleev NK, Fernandez DCDR, Lou J, Carpenter MH. Entropy stable spectral collocation schemes for the 3-D Navier-Stokes equations on dynamic unstructured grids. *J Comput Phys* (2019) 399:108897. doi:10.1016/j.jcp.2019.108897
170. Pazner W, Persson P-O. Analysis and entropy stability of the line-based discontinuous Galerkin method. *J Sci Comput* (2019) 80:376–402. doi:10.1007/s10915-019-00942-1
171. Friedrich L, Schnücke G, Winters AR, Fernández DCDR, Gassner GJ, Carpenter MH. Entropy stable space–time discontinuous Galerkin schemes with summation-by-parts property for hyperbolic conservation laws. *J Sci Comput* (2019) 80:175–222. doi:10.1007/s10915-019-00933-2
172. Gouasmi A, Duraisamy K, Murman S (2018). On entropy stable temporal fluxes. arXiv:1807.03483.
173. Gouasmi A, Murman SM, Duraisamy K. Entropy conservative schemes and the receding flow problem. *J Sci Comput* (2019) 78:971–94. doi:10.1007/s10915-018-0793-8
174. Parsani M, Boukharfane R, Nolasco IR, Fernández DCDR, Zampini S, Hadri B, et al. High-order accurate entropy-stable discontinuous collocated Galerkin methods with the summation-by-parts property for compressible CFD frameworks: scalable SSDC algorithms and flow solver. *J Comput Phys* (2020) 424:109844. doi:10.1016/j.jcp.2020.109844
175. Ranocha H, Sayyari M, Dalcin L, Parsani M, Ketcheson DI. Relaxation Runge–Kutta methods: fully discrete explicit entropy-stable schemes for the compressible Euler and Navier–Stokes equations. *SIAM J Sci Comput* (2020) 42:A612–A638. doi:10.1137/19M1263480
176. Parsani M, Carpenter MH, Fisher TC, Nielsen EJ. Entropy stable staggered grid discontinuous spectral collocation methods of any order for the compressible Navier–Stokes equations. *SIAM J Sci Comput* (2016) 38:A3129–A3162. doi:10.1137/15m1043510
177. Ortleb S. A kinetic energy preserving DG scheme based on Gauss–Legendre points. *J Sci Comput* (2016) 71:1135–68. doi:10.1007/s10915-016-0334-2
178. Fernández DCDR. Generalized summation-by-parts operators for first and second derivatives. [PhD thesis]. University of Toronto (2015).
179. Fernández DCDR, Hicken JE, Zingg DW. Simultaneous approximation terms for multi-dimensional summation-by-parts operators. *J Sci Comput* (2017) 75:83–110. doi:10.1007/s10915-017-0523-7
180. Chan J. Entropy stable reduced order modeling of nonlinear conservation laws. *J Comput Phys* (2020) 423:109789. doi:10.1016/j.jcp.2020.109789
181. Hicken JE, Fernández DCDR, Zingg DW. Multidimensional summation-by-parts operators: general theory and application to simplex elements. *SIAM J Sci Comput* (2016) 38:A1935–A1958. doi:10.1137/15m1038360
182. Crean J, Hicken JE, Fernández DCDR, Zingg DW, Carpenter MH. Entropy-stable summation-by-parts discretization of the Euler equations on general curved elements. *J Comput Phys* (2018) 356:410–38. doi:10.1016/j.jcp.2017.12.015
183. Gassner GJ, Svärd M, Hindenlang FJ (2020).Stability issues of entropy-stable and/or split-form high-order schemes. arXiv:2007.09026
184. Ranocha H, Gassner GJ (2020). Preventing pressure oscillations does not fix local linear stability issues of entropy-based split-form high-order schemes. arXiv:2009.13139.
185. Jameson A. Formulation of kinetic energy preserving conservative schemes for gas dynamics and direct numerical simulation of one-dimensional viscous compressible flow in a shock tube using entropy and kinetic energy preserving schemes. *J Sci Comput* (2008) 34:188–208. doi:10.1007/s10915-007-9172-6
186. Kuya Y, Kawai S. A stable and non-dissipative kinetic energy and entropy preserving (KEEP) scheme for non-conforming block boundaries on Cartesian grids. *Comput Fluids* (2020) 200:104427. doi:10.1016/j.compfluid.2020.104427
187. Ranocha H. Entropy conserving and kinetic energy preserving numerical methods for the Euler equations using summation-by-parts operators. In: SJ Sherwin, D Moxey, J Peiró, PE Vincent, C Schwab, editors *Spectral and high order methods for partial differential equations ICOSAHOM 2018*. Cham: Springer International Publishing (2020). p. 525–35.
188. Pirozzoli S. Numerical methods for high-speed flows. *Annu Rev Fluid Mech* (2011) 43:163–94. doi:10.1146/annurev-fluid-122109-160718
189. Dalcin L, Rojas D, Zampini S, Fernández DCDR, Carpenter MH, Parsani M. Conservative and entropy stable solid wall boundary conditions for the compressible Navier–Stokes equations: adiabatic wall and heat entropy transfer. *J Comput Phys* (2019) 397:108775. doi:10.1016/j.jcp.2019.06.051
190. Dubois F, LeFloch P. Boundary conditions for nonlinear hyperbolic systems of conservation laws. *J Differ Equ* (1988) 71:93–122. doi:10.1016/0022-0396(88)90040-X
191. Hindenlang FJ, Gassner GJ, Kopriva DA. Stability of wall boundary condition procedures for discontinuous Galerkin spectral element approximations of the compressible Euler equations. In: SJ Sherwin, D Moxey, J Peiró, PE Vincent, C Schwab, editors. *Spectral and high order methods for partial differential equations ICOSAHOM 2018*. Springer International Publishing (2020). p. 3–19.
192. Parsani M, Carpenter MH, Nielsen EJ. Entropy stable wall boundary conditions for the three-dimensional compressible Navier–Stokes equations. *J Comput Phys* (2015c) 292:88–113. doi:10.1016/j.jcp.2015.03.026
193. Svärd M. Entropy stable boundary conditions for the Euler equations. *J Comput Phys* (2020) 109947. doi:10.1016/j.jcp.2020.109947
194. Svärd M, Mishra S. Entropy stable schemes for initial-boundary-value conservation laws. *Z Angew Math Phys* (2012) 63:985–1003. doi:10.1007/s00033-012-0216-x
195. Svärd M, Özcan H. Entropy-stable schemes for the Euler equations with far-field and wall boundary conditions. *J Sci Comput* (2014) 58:61–89. doi:10.1007/s10915-013-9727-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

INTERDISCIPLINARY PHYSICS

Michael Dumbser obtained his PhD in 2004 under the supervision of Prof. Claus-Dieter Munz at Stuttgart University, Germany. Following that, he went to Trento University in Italy as a postdoc, where Prof. Tito Toro was his postdoc advisor. Already two years later in 2007 he was made a professor there, at first assistant, then associate and now full professor. One of his many honors is that he is a recipient of an ERC starting grant.

This overview article describes a class of up-to-date numerical methods that solve continuum equations that model both elastic bodies and fluid motion. The innovative aspects here are both the high fidelity numerical methods and also the model of partial differential equations describing continuum mechanics models describing a wide range of physical phenomena. The article will prove to be very useful for practitioners.

frontiers in Physics

ORIGINAL RESEARCH
published: 12 March 2020
doi: 10.3389/fphy.2020.00032

High Order ADER Schemes for Continuum Mechanics

Saray Busto, Simone Chiocchetti, Michael Dumbser, Elena Gaburro and Ilya Peshkov*

Laboratory of Applied Mathematics, Department of Civil, Environmental and Mechanical Engineering, University of Trento, Trento, Italy

In this paper we first review the development of high order ADER finite volume and ADER discontinuous Galerkin schemes on fixed and moving meshes, since their introduction in 1999 by Toro et al. We show the modern variant of ADER based on a space-time predictor-corrector formulation in the context of ADER discontinuous Galerkin schemes with a posteriori subcell finite volume limiter on fixed and moving grids, as well as on space-time adaptive Cartesian AMR meshes. We then present and discuss the unified symmetric hyperbolic and thermodynamically compatible (SHTC) formulation of continuum mechanics developed by Godunov, Peshkov, and Romenski (GPR model), which allows to describe fluid and solid mechanics in one single and unified first order hyperbolic system. In order to deal with free surface and moving boundary problems, a simple diffuse interface approach is employed, which is compatible with Eulerian schemes on fixed grids as well as direct Arbitrary-Lagrangian-Eulerian methods on moving meshes. We show some examples of moving boundary problems in fluid and solid mechanics.

Keywords: Godunov-Peshkov-Romenski model, high order, finite volume, discontinuous Galerkin, diffuse interface

OPEN ACCESS

Edited by:
Christian F. Klingenberg, Julius Maximilian University of Wrzburg, Germany

Reviewed by:
Igor Petrov, Moscow Institute of Physics and Technology, Russia
Anna Carbone, Politecnico di Torino, Italy

****Correspondence:***
Michael Dumbser
michael.dumbser@unitn.it

Specialty section:
This article was submitted to Interdisciplinary Physics, a section of the journal Frontiers in Physics

Received: *04 December 2019*
Accepted: *04 February 2020*
Published: *12 March 2020*

Citation:
Busto S, Chiocchetti S, Dumbser M, Gaburro E and Peshkov I (2020) High Order ADER Schemes for Continuum Mechanics. Front. Phys. 8:32. doi: 10.3389/fphy.2020.00032

1. INTRODUCTION AND REVIEW OF THE ADER APPROACH

The development of high order numerical schemes for hyperbolic conservation laws has been one of the major challenges of numerical analysis for the last decades. Godunov [1] proved that for the linear advection equation no monotone linear schemes of second or higher order of accuracy can be constructed. Therefore, even if physical viscosity is considered, a linear high order scheme will present spurious oscillations near discontinuities, as it can be seen, for instance for the Lax-Wendroff scheme, Lax and Wendroff [2]. A first idea to circumvent this theorem has been proposed in Kolgan [3], where limited slopes are employed to produce a non-linear scheme of second order of accuracy in space. Since then, many high order numerical methods have been developed like the Total Variation Disminishing methods (TVD) and Flux limiter methods (see, for instance, [4–9]). Despite these methodologies being already well-established at the end of the last century, their major drawback was that they just provided global second order of accuracy and reduced locally to first order in the vicinity of smooth extrema.

More advanced non-linear methods for advection dominated problems involve the family of ENO and WENO schemes, see Harten and Osher [10], Harten et al. [11], and Shu [12]. In particular, the method of Harten et al. [11] is a fully discrete high order scheme that can be re-interpreted in terms of the solution of a generalized Riemann problem (GRP), see Castro and Toro [13]. Moreover, it can be seen as a generalization of the MUSCL-Hancock method of van Leer, see van Leer [8], Toro [9], and Berthon [14].

Following the idea of solving a generalized Riemann problem (GRP), see also Ben-Artzi and Falcovitz [15], LeFloch and Tatsien [16], Ben-Artzi et al. [17], and Han et al. [18], the ADER approach (Arbitrary high order DErivative Riemann problem) has been first put forward for the linear advection equation with constant coefficients by Millington et al. [19] and Toro et al. [20]. The first step of the methodology involves piece-wise polynomial data reconstruction, where a non-linear ENO reconstruction is applied in order to avoid spurious oscillations of the numerical solution. Then, a GRP is defined at each cell interface. Classically, the initial condition for the GRP was given as piece-wise linear polynomials and second order schemes could be obtained by constructing a space-time integral of the solution in an appropriate control volume [21, 22], or following a MUSCL approach, van Leer [23] and Colella [24]. An alternative methodology proposed in Ben-Artzi and Falcovitz [25] consists in expressing the solution of the GRP as a Taylor series expansion in time. The ADER approach obtains the high order time derivatives of the GRP solution at the cell interface via the Cauchy-Kovalevskaya procedure, which replaces time derivatives by spatial derivatives using repeated differentiation of the differential form of the PDE. The spatial derivatives, which may also jump at the interface, are defined via the solution of *linearized* Riemann problems for the derivatives, where linearization is carried out about the Godunov state obtained from the classical Riemann problem between the boundary extrapolated values at the interface. In **Figure 1**, the classical piece-wise constant polynomials are plotted against a high order reconstruction and the similarity solutions for both cases are sketched. Finally, these similarity solutions are used to construct the numerical flux. The resulting schemes are arbitrary high order accurate in both space and time, in the sense that they have no theoretical accuracy barrier.

Since their introduction in Toro et al. [20] and Millington et al. [19], many extensions of the ADER methodology have been proposed. Regarding 2D linear PDEs, one may refer to Schwartzkopff et al. [26] and their simplification for the particular case of structured grids in Schwartzkopff et al. [27]. Moreover, non-linear systems have been initially addressed in Toro and Titarev [28] and Titarev and Toro [29]. Further applications of ADER on non-Cartesian meshes have been presented in Käser [30], Käser and Iske [31], Dumbser et al. [32], and Castro and Toro [13]. One should also mention the development of ADER schemes in the framework of discontinuous Galerkin (DG) finite element methods, see Qiu et al. [33], Dumbser and Munz [34] and Gassner et al. [35]. One of the main advantages of using DG is that the reconstruction step of classical ADER finite volume (ADER-FV) schemes can be skipped, since the discrete solution is already given by high order piecewise polynomials that can be directly evolved during each time step. Furthermore, ADER-DG schemes avoid the use of classical Runge-Kutta time stepping and thus provide efficient communication-avoiding schemes for parallel computing, see Fambri et al. [36] and allow for simple and natural time-accurate local time stepping (LTS), see Dumbser et al. [37].

An important step forward in the development of more general ADER schemes was achieved in Dumbser et al. [38], where a new class of ADER-FV methods has been introduced. The main contribution of this paper consists in the introduction of a new element-local space-time DG predictor, which allows at the same time the treatment of stiff source terms, as well as the replacement of the cumbersome Cauchy-Kovalevskaya procedure. First, a high order WENO method is employed to compute a polynomial reconstruction of the data inside each spatial element; then, an element-local weak formulation of the conservation law is considered in space-time and the predictor is applied to construct the time evolution of the WENO polynomials within each cell. Note that, in this step, the integration by parts is performed only in time, which differs from global space-time DG schemes [39, 40], which are globally implicit. Finally, the cell averages are updated with an explicit fully discrete one-step scheme, considering the integral form of the equations. As a result, the proposed methodology maintains arbitrary high order of accuracy, while avoiding the issues related to the use of a Taylor series expansion in time. As already mentioned above, it naturally provides an approach for the treatment of stiff source terms [for further details on this topic, see [41] and references therein].

The above methodology can also be applied in the discontiuous Galerkin framework as presented in Dumbser et al. [42], where, a unified $P_N P_M$ framework for arbitrary high order one-step finite volume and DG schemes has been introduced. For other reconstruction-based DG schemes, see e.g., Luo et al. [43, 44]. Afterwards, the methodology has been extended to solve a wide variety of different PDE systems, such as the resistive relativistic MHD equations, Dumbser and Zanotti [45]; non-conservative hyperbolic systems found in geophysical flows, Dumbser et al. [46] in which a well-balanced and path-conservative version of the scheme has been developed; compressible multi-phase flows Dumbser et al. [47], the compressible Navier-Stokes equations, Dumbser [48]; the compressible Euler equations and divergence-free schemes for MHD, Balsara et al. [49], and Balsara and Dumbser [50], where ADER schemes were used in combination with genuinely multidimensional Riemann solvers. The last extensions concern the special and general relativistic MHD equations, see Zanotti et al. [51], and Fambri et al. [36], as well as the Einstein field equations of general relativity [52, 53].

Later, ADER schemes have been extended to adaptive mesh refinement on Cartesian grids (AMR), in combination with time accurate local time stepping (LTS). This technique has initially been introduced in Dumbser et al. [54, 55] for conservative and non-conservative hyperbolic systems, respectively. Moreover, the schemes of the ADER family were the first high order methods to be applied for the numerical solution of the unified first order hyperbolic formulation of continuum mechanics by Godunov, Peshkov and Romenski [56–58], see Dumbser et al. [59–61]. In the rest of this paper, we will refer to the Godunov-Peshkov-Romenski model of continuum mechanics as GPR model.

The ADER approach has also been extended to the direct Arbitrary-Lagrangian-Eulerian framework (ALE), where the mesh moves with an arbitrary velocity, taken as close as possible to the local fluid velocity. Initially developed for one space dimension, it has been soon extended to the case of the two

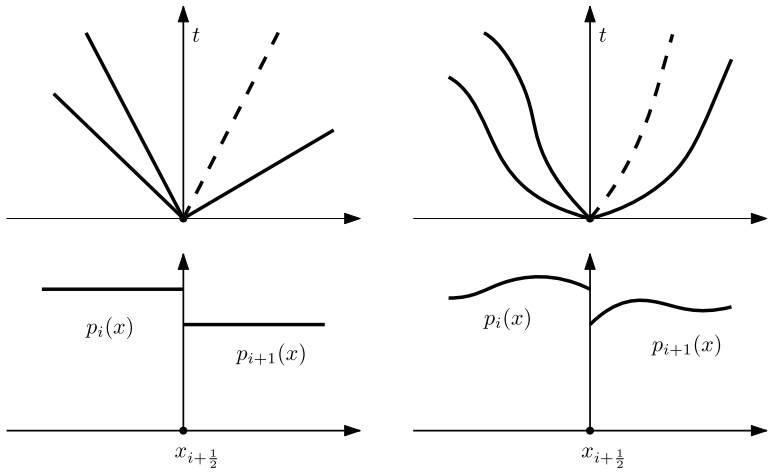


FIGURE 1 | Classical piece-wise reconstruction polynomials used in the ADER approach, $p_i(x)$ and $p_{i+1}(x)$, and the structure of the Riemann problem solution at the intercell boundary $x_{i+\frac{1}{2}}$. **(Left)** classical piece-wise constant data. **(Right)** piece-wise smooth reconstruction.

and three dimensional Euler equations on unstructured meshes, Boscheri and Dumbser [62, 63], including the discretization of non-conservative products. Further works in this area involve the use of local timestepping techniques, [64, 65]; coupling with multidimensional HLL Riemann solvers, Boscheri et al. [66]; solution of magnetohydrodynamics problems (MHD), [67, 68]; development of a quadrature-free approach to increase the computational efficiency of the overall method, Boscheri and Dumbser [69]; use of curvilinear unstructured meshes, Boscheri and Dumbser [70]; or extension to solve the GPR model, Boscheri et al. [71] and Peshkov et al. [72]. Furthermore, in Gaburro et al. [73] a novel algorithm to deal with moving non-conforming polygonal grids has been presented. The methodology reduces the typical mesh distortion arising in shear flows and provides high quality elements even for long-time simulations. An exactly well-balanced path-conservative version of this approach for the Euler equations with gravity can be found in Gaburro et al. [74]. Still in the ALE framework, within this article, we will present new results for the family of ADER-FV and ADER-DG schemes on moving unstructured Voronoi meshes [75], as recently introduced in Gaburro [76] and Gaburro et al. [77].

It is well-known that when dealing with high order schemes special care must be paid to the limiting methodology employed. In most of the previous referenced papers classical *a priori* limiters have been used, such as WENO reconstruction. Nevertheless, some alternative contributions to this topic can be found in the series of papers [51, 77–85], where a novel *a posteriori* sub-cell FV limiter of high order DG schemes, based on the MOOD paradigm of Clain et al. [86] and Diot et al. [87, 88], has been employed.

Besides the references given above, which focus on the development of the ADER methodology with a local space-time Galerkin predictor, many recent papers have been devoted to the development of other families of ADER schemes, like the classical ADER finite volume methods. Without pretending to be exhaustive, we may refer to Castro et al. [89], Toro and Hidalgo [90], Taube et al. [91], Toro [9], Montecinos et al. [92], Montecinos and Toro [93], Toro and Montecinos [94, 95], Toro et al. [96], Busto [97], Montecinos et al. [98], Busto et al. [99], Contarino et al. [100], Busto et al. [101], and Dematté et al. [102] and references therein.

In this paper, as a promising application of the family of ADER schemes, we solve a diffuse interface formulation of the GPR model of continuum mechanics. In comparison with existing continuum mechanics models, the novel feature of the GPR model is in that it incorporates the two main branches of continuum mechanics, fluid and solid mechanics, in one single unified PDE system. Recall that traditionally fluid and solid mechanics are described by PDE systems of different types, i.e., parabolic (viscous fluids) and hyperbolic (linear elasticity and hyperelasticity), which imposes many theoretical and technical difficulties if one wishes to model natural and industrial processes involving co-existence of the fluid and solid states such as in fluid-structure interaction (FSI) problems, modeling of general solid-fluid transition such as in melting and solidification processes, e.g., additive manufacturing, see for example Francois et al. [103], flows of granular media [104], viscoplastic flows, e.g., debris flows, avalanches, mantle convection, flows of many industrial Bingham-type fluids, see Balmforth et al. [105]. Due to the unified treatment of fluids and solids, the GPR model thus has a great potential for simplifying the modeling process and code development for solving the aforementioned problems. Yet, before to be applied to practical problems, the GPR model may require a coupling with an interface tracking/capturing technique for the modeling of moving material boundaries such as in free surface flows or solid body motion. In particular, in this paper, we couple the GPR model with a simple diffuse interface approach, see Tavelli et al. [85], Dumbser [106], Gaburro et al. [107], Kemm et al. [108]. For example, very interesting computational results with similar diffuse interface approaches and level set techniques for compressible multi-material flows have been obtained for example in Gavrilyuk et al. [109], Favrie et al. [110], Favrie and

Gavrilyuk [111], Ndanou et al. [112], de Brauer et al. [113], Michael and Nikiforakis [114], Jackson and Nikiforakis [115], and Barton [116]. Finally, we demonstrate that the ADER family of schemes is capable to resolve the GPR model in both solid and fluid regimes.

The paper is organized as follows. In section 2 we present the family of ADER finite volume and ADER discontinuous Galerkin finite element schemes on fixed Cartesian and moving polygonal meshes in two space dimensions. Next, in section 3 we introduce the diffuse interface formulation of the GPR model. In section 4 we show some computational results obtained with different kinds of ADER schemes (ADER-FV and ADER-DG) on different mesh topologies, including moving unstructured Voronoi meshes, as well as fixed and adaptive Cartesian grids. The paper is rounded off by some concluding remarks and an outlook to future work in section 5.

2. ADER FINITE VOLUME AND DISCONTINUOUS GALERKIN SCHEMES

The numerical method adopted in this paper is the variant of the arbitrary high-order accurate ADER approach based on the space-time predictor-corrector formalism, which we have briefly reviewed in the previous section 1. It easily applies to the context of finite volume (FV) and discontinous Galerkin (DG) methods, using either space-time adaptive Cartesian grids (AMR), see Bungartz et al. [117],Weinzierl and Mehl [118], Dumbser et al. [54], Zanotti et al. [80], Fambri et al. [36, 84] and references therein, or unstructured meshes, and both on fixed Eulerian domains or in a moving Arbitrary-Lagrangian-Eulerian (ALE) framework, see Boscheri et al. [65, 68], Boscheri and Dumbser [62, 63, 119], Boscheri [82], Gaburro [120], Gaburro et al. [77], and references therein.

Here, we briefly describe the key features of our numerical scheme, keeping the notation as general as possible, and referring to the literature for further details. We start by introducing the general form of our governing PDE system and a moving unstructured discretization of two-dimensional domains (sections 2.1 and 2.2); next, in section 2.3 we describe the data representation of the discrete solution. Then, we explain how to obtain high order of accuracy in *space*: this is available by construction in the DG case, and obtained via some variants of the well-known WENO procedure [32, 121–125] for the FV approach. Finally, we focus on the predictor-corrector version of the ADER scheme that allows to achieve arbitrary high order of accuracy in space and *time*. Since it is out of the scope of this paper to recall all the details, a general overview is given in sections 2.5 and 2.7, and an inedited proof of the convergence of the predictor for a non-linear conservation law is presented in section 2.6.

We would like to emphasize that, besides this novel convergence proof, other progress has been introduced within this work. Indeed, up to our knowledge, it is the first time that: (i) the ADER approach is used to solve a diffuse interface formulation of the GPR model that addresses the free surface problem in both solid and fluid mechanics context (previously, a similar formulation was used only in the solid dynamics context [112, 126, 127]); (ii) non-conservative products are taken into account in the high order direct ALE scheme of Gaburro et al. [77], where they have to be integrated also on degenerate space–time control volumes (see section 2.5.2).

2.1. Governing PDE System

In this paper we consider high order fully-discrete schemes for non-linear systems of hyperbolic PDE with non-conservative products and algebraic source terms of the form

$$\frac{\partial \mathbf{Q}}{\partial t} + \nabla \cdot \mathbf{F}(\mathbf{Q}) + \mathbf{B}(\mathbf{Q}) \cdot \nabla \mathbf{Q} = \mathbf{S}(\mathbf{Q}), \tag{1}$$

where $\mathbf{Q} = \mathbf{Q}(\mathbf{x}, t) \in \Omega_Q \subset \mathbb{R}^m$ is the state vector, $t \in \mathbb{R}_0^+$ is the time, $\mathbf{x} \in \Omega \subset \mathbb{R}^d$ is the spatial coordinate, d is the number of space dimensions, Ω_Q is the so-called state space or phase space, $\mathbf{F}(\mathbf{Q})$ is the non-linear flux tensor, $\mathbf{B}(\mathbf{Q})\cdot\nabla\mathbf{Q}$ is a non-conservative product and $\mathbf{S}(\mathbf{Q})$ is a purely algebraic source term. Introducing the system matrix $\mathbf{A}(\mathbf{Q}) = \partial\mathbf{F}/\partial\mathbf{Q} + \mathbf{B}(\mathbf{Q})$ the above system can also be written in quasi-linear form as

$$\frac{\partial \mathbf{Q}}{\partial t} + \mathbf{A}(\mathbf{Q}) \cdot \nabla \mathbf{Q} = \mathbf{S}(\mathbf{Q}). \tag{2}$$

The system is said to be hyperbolic if for all $\mathbf{n} \neq 0$ and for all $\mathbf{Q} \in \Omega_Q$ the matrix $\mathbf{A}(\mathbf{Q}) \cdot \mathbf{n}$ has m real eigenvalues and a full set of m linearly independent right eigenvectors. The system (1) needs to be provided with an initial condition $\mathbf{Q}(\mathbf{x}, 0) = \mathbf{Q}_0(\mathbf{x})$ and appropriate boundary conditions on $\partial\,\Omega$.

In this paper we focus on a particular, but very general, example of a first-order system (1) describing elastic and visco-plastic heat-conducting media; it will be discussed in section 3.

2.2. Domain Discretization

In the general ALE case, we consider a moving two-dimensional ($d = 2$) domain $\Omega(t)$ and we cover it using an unstructured mesh made of N_P non-overlapping polygons $P_i, i = 1, \ldots N_P$. The mesh is first built at time $t = 0$ and then it is rearranged at each time step t^n: elements and nodes are moved following the local fluid velocity and when necessary, in order to prevent mesh distortion, also the mesh topology (i.e., the shape of the elements and their connectivities) is changed.

Given a polygon P_i^n we denote by $\mathcal{V}(P_i^n) = \{v_{i_1}^n, \ldots, v_{i_j}^n, \ldots, v_{i_{N_{V_i}^n}}^n\}$ the set of its $N_{V_i}^n$ Voronoi neighbors (the neighbors that share with P_i^n at least a vertex), and by $\mathcal{E}(P_i^n) = \{e_{i_1}^n, \ldots, e_{i_j}^n, \ldots, e_{i_{N_{V_i}^n}}^n\}$ the set of its $N_{V_i}^n$ edges, and by $\mathcal{D}(P_i^n) = \{d_{i_1}^n, \ldots, d_{i_j}^n, \ldots, d_{i_{N_{V_i}^n}}^n\}$ the set of its $N_{V_i}^n$ vertexes, consistently ordered counterclockwise. Finally, the barycenter of P_i^n is noted as $\mathbf{x}_{b_i}^n = (x_{b_i}^n, y_{b_i}^n)$. When necessary, by connecting $\mathbf{x}_{b_i}^n$ with each vertex of $\mathcal{D}(P_i)$ we can subdivide a polygon P_i^n in $N_{V_i}^n$ subtriangles denoted as $\mathcal{T}(P_i^n) = \{T_{i_1}^n, \ldots, T_{i_j}^n, \ldots, T_{i_{N_{V_i}^n}}^n\}$.

The coordinates of each node at time t^n are denoted by $\mathbf{x}_k^n$, and $\overline{\mathbf{V}}_k^n$ represents the velocity at which it is supposed to move,

so that its new coordinates at time t^{n+1} are given from the following relation

$$\mathbf{x}_k^{n+1} = \mathbf{x}_k^n + \Delta t \overline{\mathbf{V}}_k^n. \tag{3}$$

More details on how to obtain $\overline{\mathbf{V}}$ can be found in Boscheri et al. [68], Boscheri and Dumbser [63, 119] for what concerns classical direct ALE schemes on conforming unstructured grids, in Gaburro et al. [73, 74] for non-conforming unstructured grids, in Boscheri and Dumbser [70] for curvilinear meshes, and we refer in particular to section 2.4 and 2.5 of Gaburro et al. [77] for what concerns moving unstructured polygonal grids allowing for topology changes, which indeed is the ALE case considered in this paper (see case B below). Moreover, working in the ALE framework, we are allowed to take $\overline{\mathbf{V}} = \mathbf{0}$, i.e., we can also work in a fixed *Eulerian* system where the initial mesh is never modified.

In particular, in this paper we will consider the following two situations for our domain discretization:

A. A fixed Cartesian mesh made of N_P quadrilaterals elements, which is not moved during the simulation, but which can be successively refined, with a general space-tree-type data structure that allows element-by-element refinement with a general refinement factor $\mathfrak{r} \geq 2$, in order to increase the resolution in the areas of interest, as can be seen in **Figure 2** (for the details on the refinement procedure we refer to Dumbser et al. [54] and Fambri et al. [36]). To ease the description of the numerical method, we will associate to each quadrilateral element P_i^n, a set of indices that refer to its Cartesian coordinates, $\{j,k\}$, such that $P_{jk}^n := P_i^n = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}] \times [y_{k-\frac{1}{2}}, y_{k+\frac{1}{2}}]$, $\Delta x_j = x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}$, $\Delta y_k = y_{k+\frac{1}{2}} - y_{k-\frac{1}{2}}$.
B. A moving polygonal grid as the one described in Gaburro et al. [77] that (i) moves with the fluid flow in order to reduce the numerical dissipation associated with transport terms and (ii) also allows for topology changes at any time step in order to maintain always a high quality of the moving mesh; in this case we remark that our method is also able to deal with degenerate space time control volumes at arbitrary high order of accuracy.

2.2.1. Space-Time Connectivity

To better understand the context of moving meshes we refer the reader to **Figure 3**: note that the tessellation at time t^n has been evolved resulting in a slightly different tessellation at time t^{n+1}; for each element P_i^n the new vertex coordinates $\mathbf{x}_k^{n+1}$, $k = 1, \ldots, N_{V_i}^n$, are connected to the old coordinates $\mathbf{x}_k^{n+1}$ via straight line segments, yielding the multidimensional *space-time control volume* C_i^n, that involves $N_{V_i}^{n,st}+2$ space-time sub-surfaces. Specifically, the space-time volume C_i^n is bounded on the bottom and on the top by the element configuration at the current time level P_i^n and at the new time level P_i^{n+1}, respectively, while it is closed with a total number of $N_{V_i}^{n,st}$ lateral space–time surfaces $\partial C_{ij}^n, j = 1, \ldots, N_{V_i}^{n,st}$ that are given by the evolution of each edge e_{ij}^n of element P_i^n within the time step $\Delta t = t^{n+1} - t^n$. *A priori*, ∂C_{ij}^n are not parallel to the time direction: thus to be treated numerically they can be mapped to a reference square by using a set of of bilinear basis functions (see Boscheri and Dumbser [62]). To resume, the space-time volume C_i^n is bounded by its surface ∂C_i^n which is given by

$$\partial C_i^n = \left(\bigcup_j \partial C_{ij}^n \right) \cup P_i^n \cup P_i^{n+1}. \tag{4}$$

Note that in the fixed Cartesian case, C_i^n reduces to a right parallelepiped with four lateral space–time surfaces ∂C_{ij}^n parallel to the time-direction, so many simplifications are possible.

We close this part by emphasizing that the family of direct ALE schemes proposed in this work, based on the ADER predictor-corrector approach, is based on the integration of the governing Equation (1) *in space and in time* directly over these *space–time* control volumes, see section 2.7. Note that this procedure, which is more evident when C_i^n is an oblique prism, is also hidden when C_i^n is just a right parallelepiped.

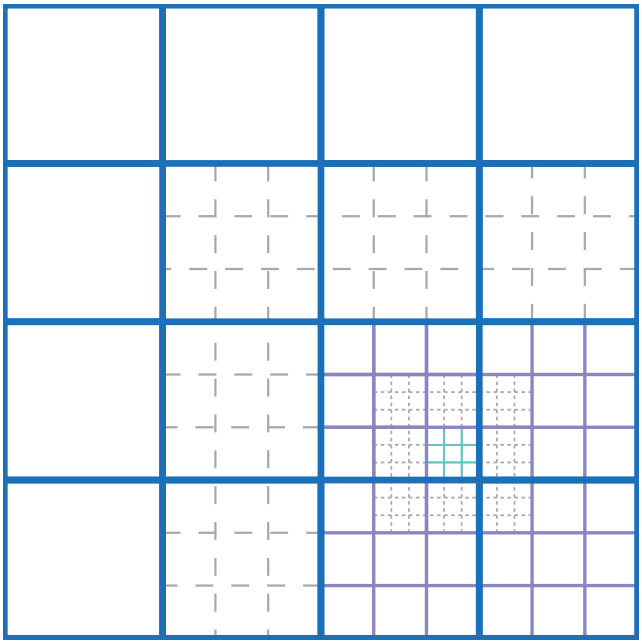


FIGURE 2 | Sketch of the mesh refinement structure of three AMR levels with refinement factor $\mathfrak{r} = 3$. Solid lines indicate active cells, whereas the dashed ones are the virtual cells allowing interpolation between the coarse and the refined mesh, needed in the case of high order WENO reconstruction.

2.3. Data Representation

The conserved variables $\mathbf{Q}$ in (1) are discretized in each polygon P_i^n at the current time t^n via piecewise polynomials of arbitrary high order N, denoted by $\mathbf{u}_h^n(\mathbf{x}, t^n)$ and defined as

$$\mathbf{u}_h^n(\mathbf{x}, t^n) = \sum_{\ell=0}^{\mathcal{N}-1} \varphi_\ell(\mathbf{x}, t^n)\, \hat{\mathbf{u}}_{\ell,i}^n = \varphi_\ell(\mathbf{x}, t^n)\, \hat{\mathbf{u}}_{\ell,i}^n, \quad \mathbf{x} \in P_i^n, \tag{5}$$

where in the last equality we have employed the classical tensor index notation based on the Einstein summation convention,

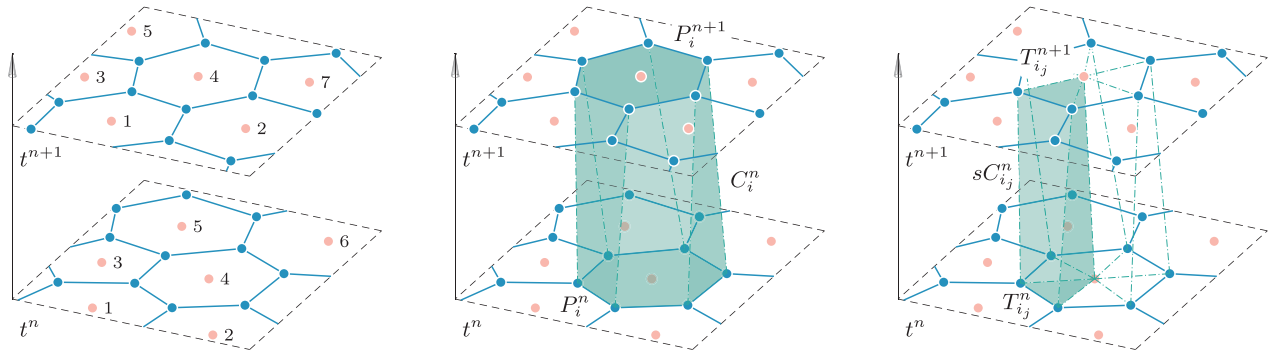


FIGURE 3 | Space time connectivity. **(Left)** The tessellation at time t^n and time t^{n+1}. **(Middle)** P_i^n is connected with P_i^{n+1} to construct the space–time control volume C_i^n. **(Right)** The sub-triangle T_{ij}^n is connected with T_{ij}^{n+1} to construct the sub–space–time control volume sC_{ij}^n.

which implies summation over two equal indices. The functions $\varphi_\ell(\mathbf{x}, t^n)$ can be either:

i. *Nodal* spatial basis functions given by a set of Lagrange interpolation polynomials of maximum degree N with the property

$$\varphi_\ell(\mathbf{x}_{\mathrm{GL}}^m) = \begin{cases} 1 \text{ if } \ell = m; \\ 0 \text{ otherwise;} \end{cases} \qquad \ell, m = 1, \ldots, (N+1)^{\mathrm{d}}, \quad (6)$$

where $\{\mathbf{x}_{\mathrm{GL}}^m\}$ are the set of the Gauss-Legendre (GL) quadrature points on P_i^n (see Stroud [128] for the multidimensional case). In particular, when employing these basis functions on a Cartesian grid, each quadrilateral P_i^n is easily mapped to a reference square, we only need the tensor product of the GL quadrature points in the unit interval $[0, 1]$, and the φ_ℓ are simply generated by multiplying one-dimensional nodal basis functions, i.e.,

$$\varphi_\ell(\mathbf{x}, t^n) = \varphi_{\ell_1}\left(\xi(x)\right) \varphi_{\ell_2}\left(\eta(y)\right) \quad (7)$$

with φ_{ℓ_i} satisfying (6) with $d = 1$, and $x = x_{j-\frac{1}{2}} + \xi \Delta x_j$, $y = y_{k-\frac{1}{2}} + \eta \Delta y_k$ being the set of reference coordinates related to P_i^n. In this case, the total number of GL quadrature points per polygon, as well as the total number of basis functions $\{\varphi_\ell\}$ and expansion coefficients $\hat{\mathbf{u}}_{\ell,i}^n$, the so-called degrees of freedom (DOF), is $\mathcal{N} = (N+1)^d$. These basis functions are used on Cartesian grids, i.e., for Case A.

ii. *Modal* spatial basis functions written through a Taylor series of degree N in the variables $\mathbf{x} = (x, y)$ directly defined on the *physical element* P_i^n, expanded about its current barycenter $\mathbf{x}_{b_i}^n$ and normalized by its current characteristic length h_i

$$\varphi_\ell(\mathbf{x}, t^n)|_{P_i^n} = \frac{(x - x_{b_i}^n)^{p_\ell}}{p_\ell!\, h_i^{p_\ell}} \frac{(y - y_{b_i}^n)^{q_\ell}}{q_\ell!\, h_i^{q_\ell}},$$
$$\ell = 0, \ldots, \mathcal{N} - 1, \quad 0 \le p_\ell + q_\ell \le N, \quad (8)$$

h_i being the radius of the circumcircle of P_i^n. In this case the total number $\mathcal{N}$ of DOF $\hat{\mathbf{u}}_l^n$ is $\mathcal{N} = \frac{1}{d!} \prod_{m=1}^{d} (N + m)$. We employ this kind of basis functions in the moving unstructured polygonal Case B.

The discontinuous finite element data representation (5) leads naturally to discontinuous Galerkin (DG) schemes if $N > 0$, but also to finite volume (FV) schemes in the case $N = 0$. This indeed means that for $N = 0$ we have $\varphi_\ell(\mathbf{x}) = 1$, with $\ell = 0$ and (5) reduces to the classical piecewise constant data that are typical of finite volume methods. In the case $N > 0$ (DG) the form given by (5) already provides a spatially high order accurate data representation with accuracy $N + 1$, where instead for the case $N = 0$ (FV), if we are interested in increasing the spatial order of accuracy, up to $M + 1$ for examle, we need to perform a spatial *reconstruction*. With this notation, our method falls within the more general class of $P_N P_M$ schemes introduced in Dumbser et al. [42] for fixed unstructured meshes.

2.4. Data Reconstruction

In this section we focus on the reconstruction procedure needed in the finite volume context ($N = 0, M > 0$) in order to obtain order of accuracy $M + 1$ in space starting from the piecewise constant values of $\mathbf{u}_h^n(\mathbf{x}, t^n)$ in P_i^n and its neighbors, i.e., in order to obtain a high order polynomial of degree M representing our solution in each P_i^n

$$\mathbf{w}_h^n(\mathbf{x}, t^n) = \sum_{\ell=0}^{\mathcal{M}-1} \psi_\ell(\mathbf{x}, t^n)\, \hat{\mathbf{w}}_{\ell,i}^n = \psi_\ell(\mathbf{x}, t^n)\, \hat{\mathbf{w}}_{\ell,i}^n, \mathbf{x} \in P_i^n, \quad (9)$$

where the ψ_ℓ functions simply coincide with the φ_ℓ basis functions of (5). Our reconstruction procedures are based on the WENO algorithm in its *polynomial* formulation as presented in Dumbser et al. [38], Dumbser and Käser [32, 123], Titarev et al. [129], Tsoutsanis et al. [130], Levy et al. [131], Dumbser et al. [132], and Semplice et al. [133], and not based on the original version of WENO proposed in Jiang and Shu [121], Balsara and Shu [122], Hu and Shu [134], and Zhang and Shu [124] which provides only *point values*. For each P_i^n, the basic idea consists in

(i) selecting a central stencil of elements $\mathcal{S}_i^0$ with a total number of

$$n_e = f \cdot \frac{1}{d!} \prod_{m=1}^{d} (M+m) \tag{10}$$

elements, containing the cell P_i^n itself, its first layer of Voronoi neighbors $\mathcal{V}(P_i^n)$ and filled by recursively adding neighbors of those elements that have been already included in the stencil, and in (ii) using the cell-average values of the elements of $\mathcal{S}_i^0$ to reconstruct a polynomial of degree M by imposing the integral conservation criterion, i.e., by requiring that its average on each cell match the known cell average. If $f > 1$ (which occurs in the unstructured case, where we take $f = 1.5$), this of course leads to an overdetermined linear system, which is solved using a constrained least-squares technique (CLSQ) [123], i.e., the reconstructed polynomial has exactly the cell average $\hat{\mathbf{u}}_{0,i}^n$ on the polygon P_i^n and matches all the other cell averages of the remaining stencil elements in the least-square sense.

However, as well-known thanks to the Godunov theorem [1], the use of only one central stencil (which is indeed a linear procedure) would introduce oscillations in the presence of shock waves or other discontinuities. So, in order to make the reconstruction procedure non-linear, we will compute the final reconstruction polynomial as a *non-linear combination* or *more* than only one reconstruction polynomial, each one defined on a different reconstruction stencil $\mathcal{S}_i^s$.

We refer to the cited literature for further details, and here we just highlight the main characteristics of the two reconstruction procedures adopted in this work.

2.4.1. Case A: Cartesian Mesh

In Case A, of a fixed Cartesian mesh, we employ the polynomial WENO procedure given in Dumbser et al. [54], which is implemented in a dimension by dimension fashion. For each cell, we define its related sets of one-dimensional reconstruction stencils as

$$\mathcal{S}_i^{s,x} = \bigcup_{m=j-L}^{j+R} P_{mk}^n, \quad \mathcal{S}_i^{s,y} = \bigcup_{m=k-L}^{k+R} P_{jm}^n, \tag{11}$$

where $L = \{M, s\}$ and $R = \{M, s\}$ denote the order and stencil dependent spatial extension of the stencil to the left and to the right. For odd order schemes we consider three stencils, one central, one fully left–sided, and one fully right–sided stencil in each space dimension (see **Figure 4** for a graphical interpretation for $M = 2$), while for even order schemes we have four stencils, two of which are central, while the remaining two are again given by the fully left–sided and fully right–sided in each space dimension. In both cases the total amount of elements in each stencil is always $n_e = M + 1$, the order of the scheme.

Focusing on the reconstruction procedure on the x direction, given a element P_i^n, we start by expressing the first coordinate of the reconstruction polynomial at each stencil in terms of one dimensional basis functions,

$$\mathbf{w}_h^{s,x}(x, t^n) = \sum_{\ell_1=0}^{M} \psi_{\ell_1}(\xi)\, \hat{\mathbf{w}}_{jk,\ell_1}^{n,s} = \psi_{\ell_1}(\xi)\, \hat{\mathbf{w}}_{jk,\ell_1}^{n,s}. \tag{12}$$

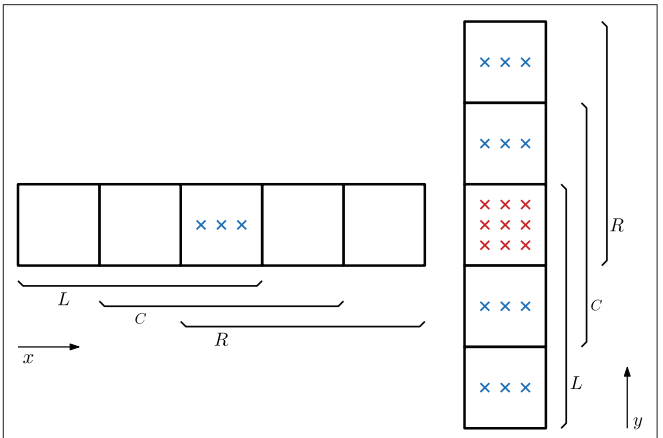


FIGURE 4 | Reconstruction stencils for a fixed Cartesian mesh with $M = 2$, where L, C, and R denote the left–sided, central and right–sided stencils, respectively. **(Left)** Reconstruction on x direction. **(Right)** Reconstruction on y direction.

Then, we integrate on the stencil elements obtaining an algebraic system on the polynomial coefficients:

$$\frac{1}{\Delta x_m} \int_{x_{m-\frac{1}{2}}}^{x_{m+\frac{1}{2}}} \psi_{\ell_1}\left(\xi(x)\right) \hat{\mathbf{w}}_{jk,\ell_1}^{n,s}\, dx = \bar{\mathbf{u}}_{mk}^n, \quad \forall P_{mk}^n \in \mathcal{S}_i^{s,x} \tag{13}$$

with $\bar{\mathbf{u}}_{mk}^n$ the average value obtained by integrating the solution at the previous time step on the cell P_{mk}. Once the coefficients, and thus the polynomials, related to all the stencils are obtained, we compute a reconstruction polynomial in the x direction as the data-dependent non-linear combination of these,

$$\mathbf{w}_h^x\left(x, t^n\right) = \psi_{\ell_1}(\xi)\, \hat{\mathbf{w}}_{jk,\ell_1}^n, \quad \hat{\mathbf{w}}_{jk,\ell_1}^n = \sum_{s=1}^{n_s} \omega_s \hat{\mathbf{w}}_{jk,\ell_1}^{n,s}, \tag{14}$$

where n_s is the number of stencils, $n_s = 3$ if $M = 2$ and $n_s = 4$ otherwise; and ω_s denote the non-linear weights (see Dumbser et al. [54] for further details).

To complete the reconstruction polynomial, we now repeat the above procedure in the y direction for each degree of freedom $\hat{\mathbf{w}}_{jk,\ell_1}^n$. First, we write the reconstruction polynomial in terms of the basis functions,

$$\mathbf{w}_h^{s,y}(x, y, t^n) = \psi_{\ell_1}(\xi)\, \psi_{\ell_2}(\eta)\, \hat{\mathbf{w}}_{jk,\ell_1\ell_2}^{n,s}. \tag{15}$$

Then, we solve the algebraic system

$$\frac{1}{\Delta y_m} \int_{y_{m-\frac{1}{2}}}^{y_{m+\frac{1}{2}}} \psi_{\ell_2}\left(\eta\left(y\right)\right) \hat{\mathbf{w}}_{jk,\ell_1\ell_2}^{n,s}\, dy = \hat{\mathbf{w}}_{jm,\ell_1}^n, \quad \forall P_{jm}^n \in \mathcal{S}_i^{s,y} \tag{16}$$

and calculate

$$\hat{\mathbf{w}}_{jk,\ell_1\ell_2}^n = \sum_{s=1}^{n_s} \omega_s \hat{\mathbf{w}}_{jk,\ell_1\ell_2}^{n,s}. \tag{17}$$

Finally, we get the WENO reconstruction polynomial

$$\mathbf{w}_h^n\left(\mathbf{x}, t^n\right) = \psi_{\ell_1}(\xi)\,\psi_{\ell_2}(\eta)\,\hat{\mathbf{w}}^n_{jk,\ell_1\ell_2}. \tag{18}$$

In order to enforce bounds on the WENO reconstruction polynomial, such as the condition $0 \leq \alpha \leq 1$ on the volume fraction function of for example (56a), we *rescale* the reconstruction coefficients $\hat{\mathbf{w}}^n_{jk,\ell_1\ell_2}$ around the cell average as follows:

$$\hat{\mathbf{w}}^*_{jk,\ell_1\ell_2} = \bar{\mathbf{u}}^*_{jk} + \varphi_{jk}\left(\hat{\mathbf{w}}^n_{jk,\ell_1\ell_2} - \bar{\mathbf{u}}^*_{jk}\right), \tag{19}$$

where the scaling factor φ_{jk} is computed via the Barth and Jespersen limiter (see Barth and Jespersen [135]) applied to the volume fraction function α in all Gauss-Legendre and Gauss-Lobatto quadrature nodes, i.e., $\varphi_{jk} = \min(\varphi_{jk,p})$ is the global minimum in each element, with the nodal limiter values given by

$$\varphi_{jk,p} = \begin{cases} \min\left(1, \frac{\alpha_{\max}-\bar{\alpha}}{\alpha_p-\bar{\alpha}}\right), & \text{if } \alpha_p - \bar{\alpha} > 0, \\ \min\left(1, \frac{\alpha_{\min}-\bar{\alpha}}{\alpha_p-\bar{\alpha}}\right), & \text{if } \alpha_p - \bar{\alpha} > 0, \\ 1, & \text{if } \alpha_p - \bar{\alpha} = 0. \end{cases} \tag{20}$$

Here $\alpha_{\max} = 1 - \varepsilon \leq 1$ is the upper bound of the volume fraction function and $\alpha_{\min} = \varepsilon \geq 0$ is its lower bound; $\bar{\alpha}$ denotes the cell average of α and α_p denotes the node value of α in the quadrature point $\mathbf{x}_p$ under consideration. As already mentioned above, this strategy is inspired from the Barth and Jespersen limiter [135], but also from the new bound-preserving polynomial approximation introduced in Després [136] and Campos-Pinto et al. [137]. Since the physical solution of α must satisfy $0 \leq \alpha \leq 1$, the above bound preserving limiter does *not* reduce the formal order of accuracy of the reconstruction, as proven in Després [136].

2.4.2. Case B: Moving Polygonal Mesh

In Case B of our moving and topology changing polygonal mesh we adopt a CWENO reconstruction algorithm, first introduced in Levy et al. [138–140] and Semplice et al. [133], and which can be cast in the general framework described in Cravero et al. [141]. We closely follow the work outlined in Dumbser et al. [132] and Boscheri et al. [142] for unstructured triangular and tetrahedral meshes, and extended it to moving polygonal grids in Gaburro et al. [77].

We emphasize that the main advantages of such a procedure is that only one stencil (the central one) is required to contain the total amount of elements stated in (10) and only this one is used to construct a polynomial of degree M; the other ones are used to compute polynomials of lower degree. In particular, we consider $N^n_{V_i}$ stencils S^s_i, each of them containing exactly $\hat{n}_e = (d+1)$ cells, i.e., the central cell P^n_i and two consecutive neighbors belonging to $\mathcal{V}(P^n_i)$. Refer to **Figure 5** for a graphical description of the stencils. For each stencil S^s_i we compute a linear polynomial by solving a simple reconstruction system which is not overdetermined. According to the above mentioned literature, the reconstructed polynomial obtained via a non-linear combination of the polynomial of degree M, computed over S^s_0, and of the $N^n_{V_i}$ linear polynomials, computed over S^s_i, maintains the order of convergence of the method and avoids unwanted spurious oscillations. In particular, in the case of moving meshes with topology changes, where the set of neighbors may change at any time step, the use of smaller so-called sectorial stencils significantly speeds up computations.

For the sake of uniform notation, in the DG case, i.e., when $N > 0$ and $M = N$, we trivially impose that the reconstruction polynomial is given by the DG polynomial, i.e., $\mathbf{w}^n_h(\mathbf{x}, t^n) = \mathbf{u}^n_h(\mathbf{x}, t^n)$, which automatically implies that in the case $N = M$ the reconstruction operator is simply the identity.

2.5. Space-Time Predictor Step

In this section we focus on the key feature, the element-local *space-time predictor* step, of our ADER FV-DG schemes: this part of the algorithm (the *predictor*) produces a high order approximation in both space and time of $\mathbf{Q}$ in all P^n_i. This allows to obtain a fully discrete one-step scheme that is uniformly high order accurate in both space and time.

The predictor step consists in a completely *local* procedure which solves the governing PDE (1) *in the small*, see Harten et al. [11], inside each space-time element C^n_i, and it only considers the geometry of volume C^n_i, the initial data $\mathbf{w}^n_h$ on P^n_i and the governing Equations (1), without taking into account any interaction between C^n_i and its neighbors. Because of this absence of communications, we refer to it as *local*. The procedure finally provides, for each C^n_i, a space-time polynomial data representation $\mathbf{q}^n_h$, which serves as a predictor solution, only valid inside C^n_i, to be used for evaluating the numerical fluxes, the non-conservative products and the algebraic source terms when integrating the PDE in the final *corrector* step (see section 2.7) of the ADER scheme.

The predictor $\mathbf{q}^n_h$ is a polynomial of degree M, which takes the following form

$$\mathbf{q}^n_h(\mathbf{x}, t) = \sum_{\ell=0}^{\mathcal{Q}-1} \theta_\ell(\mathbf{x}, t)\hat{\mathbf{q}}^n_\ell, \qquad (\mathbf{x}, t) \in C^n_i, \tag{21}$$

where $\theta_\ell(\mathbf{x}, t)$ can be either

i. For fixed and adaptive Cartesian grids (Case A), *nodal* space-time basis functions of degree M given by the product of one-dimensional nodal basis functions verifying (6) (with $d = 1$),

$$\theta_\ell(x, y, t) = \varphi_{\ell_1}\left(\xi(x)\right)\varphi_{\ell_2}\left(\eta(y)\right)\varphi_{\ell_3}\left(\tau(t)\right), \tag{22}$$

two of them mapped to the unit interval $[0, 1]$ as in (7) and with the time coordinate mapped to the reference time $\tau \in [0, 1]$ via $t = t^n + \tau\Delta t$. In this case, the total number of GL quadrature points per cell, as well as the total number of DOF is $\mathcal{Q} = (M+1)^{d+1}$, see also **Figure 6**.

ii. For our moving polygonal meshes (Case B), *modal* space time basis functions of degree M in $d+1$ dimensions (d space

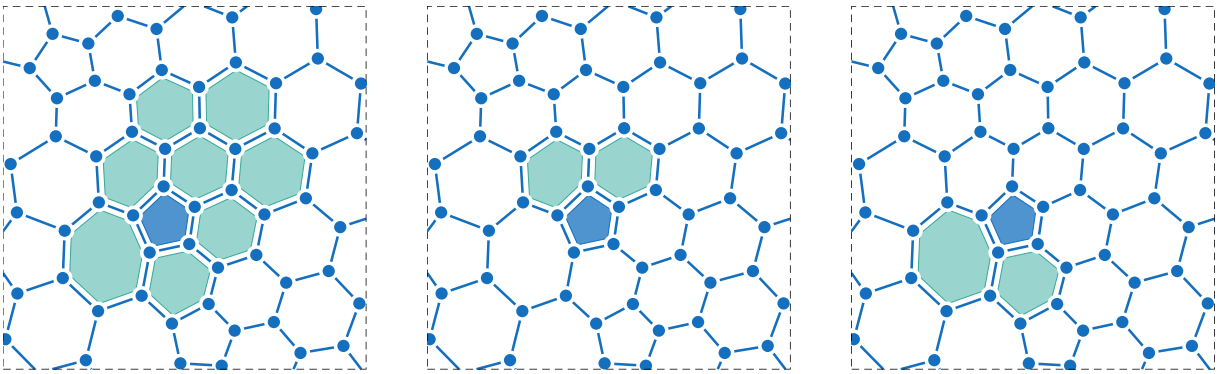


FIGURE 5 | Stencils for the CWENO reconstruction of order three ($M = 2$) with $f = 1.5$ for a pentagonal element P_i^n. Left: central stencil made of the element itself P_i^n (in violet) and $n_e - 1 = 8$ of its neighbors (in blue). In the other panels we report two of the $N_{V_i}^n = 5$ sectorial stencils containing the element itself and two consecutive neighbors belonging to $\mathcal{V}(P_i^n)$.

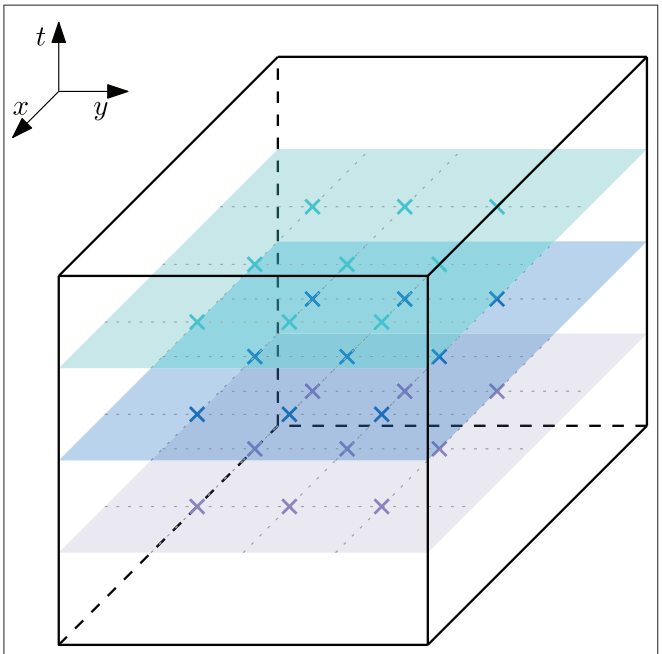


FIGURE 6 | Quadrature points on a space-time element, C_i^n, of a fixed Cartesian mesh with $M = 2$.

dimensions plus time) are used, which read

$$\theta_\ell(x,y,t)|_{C_i^n} = \frac{(x-x_{b_i}^n)^{p_\ell}}{p_\ell!\, h_i^{p_\ell}} \frac{(y-y_{b_i}^n)^{q_\ell}}{q_\ell!\, h_i^{q_\ell}} \frac{(t-t^n)^{q_\ell}}{q_\ell!\, h_i^{q_\ell}}, \quad \ell = 0,\dots,\mathcal{Q}, \quad 0 \le p_\ell + q_\ell + r_\ell \le M, \tag{23}$$

with the total number of DOF $\mathcal{Q} = \frac{1}{(d+1)!}\prod_{m=1}^{d+1}(M+m)$, see also **Figure 7**.

Now, multiplying our PDE system (1) with a test function θ_k and integrating over the space-time control volume C_i^n (see section 2.2.1), we obtain the following weak form of the governing PDE, where both the test and the basis functions are *time dependent*

$$\int_{C_i^n} \theta_k(\mathbf{x},t)\frac{\partial \mathbf{q}_h^n}{\partial t}\,d\mathbf{x}\,dt + \int_{C_i^n} \theta_k(\mathbf{x},t)\left(\nabla\cdot\mathbf{F}(\mathbf{q}_h^n) + \mathbf{B}(\mathbf{q}_h^n)\cdot\nabla\mathbf{q}_h^n\right)d\mathbf{x}\,dt = \int_{C_i^n}\theta_k(\mathbf{x},t)\mathbf{S}(\mathbf{q}_h^n)\,d\mathbf{x}\,dt\,. \tag{24}$$

Since we are only interested in an element local predictor solution, i.e., we do not need to consider the interactions with the neighbors, we do not yet take into account the jumps of $\mathbf{q}_h^n$ across the space–time lateral surfaces, because this will be done in the final corrector step (section 2.7).

Instead, we insert the known discrete solution $\mathbf{w}_h^n(\mathbf{x},t^n)$ at time t^n in order to introduce a weak initial condition for solving our PDE; note that $\mathbf{w}_h^n(\mathbf{x},t^n)$ uses information coming from the past only (following an *upwinding approach*) in such a way that the causality principle is correctly respected. To this purpose, the first term is integrated by parts in time. This leads to

$$\int_{P_i^{n+1}}\theta_k(\mathbf{x},t^{n+1})\mathbf{q}_h^n(\mathbf{x},t^{n+1})\,d\mathbf{x} - \int_{P_i^n}\theta_k(\mathbf{x},t^n)\mathbf{w}_h^n(\mathbf{x},t^n)\,d\mathbf{x} - \int_{C_i^n}\frac{\partial}{\partial t}\theta_k(\mathbf{x},t)\mathbf{q}_h^n(\mathbf{x},t)\,d\mathbf{x}\,dt + \int_{C_i^n\setminus\partial C_i^n}\theta_k(\mathbf{x},t)\nabla\cdot\mathbf{F}(\mathbf{q}_h^n)\,d\mathbf{x}\,dt = \int_{C_i^n\setminus\partial C_i^n}\theta_k(\mathbf{x},t)\left(\mathbf{S}(\mathbf{q}_h^n) - \mathbf{B}(\mathbf{q}_h^n)\cdot\nabla\mathbf{q}_h^n\right)d\mathbf{x}\,dt. \tag{25}$$

Equation (25) results in an element-local non-linear system for the unknown degrees of freedom $\hat{\mathbf{q}}_\ell^n$ of the space-time polynomials $\mathbf{q}_h^n$. The solution of (25) can be found via a simple and fast converging fixed point iteration (a discrete Picard iteration) as detailed e.g., in Dumbser et al. [42] and Hidalgo and Dumbser [41]. For linear homogeneous systems, the discrete Picard iteration converges in a finite number of at most $N+1$ steps, since the involved iteration matrix is nilpotent, see Jackson [143]. Moreover a proof of the convergence of this procedure in

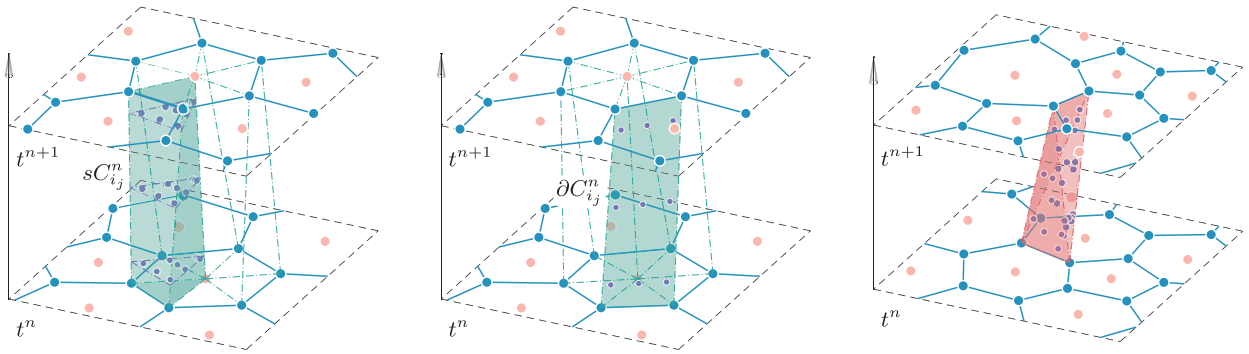


FIGURE 7 | Space–time quadrature points for third order methods, i.e., $M = 2$, on a moving polygonal mesh with topology changes. **(Left)** Quadrature points for the volume integrals and the space–time predictor. **(Middle)** quadrature points for the surface integrals, i.e., for flux computation. **(Right)** Quadrature points for the volume integrals and the space–time predictor of a sliver element.

the case of a non-linear homogeneous conservation law in 1D is given in next section 2.6.

2.5.1. Simplification in the Case of a Fixed Cartesian Mesh

The space-time predictor step formerly presented can be simplified in the case of a Cartesian mesh with nodal basis functions resulting in a more efficient algorithm. Under these assumptions the governing PDE (1), can be rewritten as

$$\frac{\partial \mathbf{Q}}{\partial \tau} + \frac{\partial \mathbf{f}^\star}{\partial \xi} + \frac{\partial \mathbf{g}^\star}{\partial \eta} + \mathbf{B}_1^\star \frac{\partial \mathbf{Q}}{\partial \xi} + \mathbf{B}_2^\star \frac{\partial \mathbf{Q}}{\partial \eta} = \mathbf{S}^\star \tag{26}$$

with

$$\mathbf{f}^\star = \frac{\Delta t}{\Delta x_j}\mathbf{f}, \quad \mathbf{g}^\star = \frac{\Delta t}{\Delta y_k}\mathbf{g}, \quad \mathbf{B}_1^\star = \frac{\Delta t}{\Delta x_j}\mathbf{B}_1, \quad \mathbf{B}_2^\star = \frac{\Delta t}{\Delta y_k}\mathbf{B}_2, \quad \mathbf{B} = [\mathbf{B}_1, \mathbf{B}_2], \quad \mathbf{S}^\star = \Delta t \mathbf{S}. \tag{27}$$

Next, we multiply each term by a test function θ_k and we integrate over the reference space-time control volume $\mathcal{I}_0 = [0,1]^3$

$$\int_0^1\int_0^1\int_0^1 \theta_k \left(\frac{\partial \mathbf{Q}}{\partial \tau} + \frac{\partial \mathbf{f}^\star(\mathbf{Q})}{\partial \xi} + \frac{\partial \mathbf{g}^\star(\mathbf{Q})}{\partial \eta} \right) d\xi d\eta d\tau = \int_0^1\int_0^1\int_0^1 \theta_k \left(\mathbf{S}^\star(\mathbf{Q}) - \mathbf{B}_1^\star(\mathbf{Q}) \frac{\partial \mathbf{Q}}{\partial \xi} - \mathbf{B}_2^\star(\mathbf{Q}) \frac{\partial \mathbf{Q}}{\partial \eta} \right) d\xi d\eta d\tau. \tag{28}$$

Now, by substituting the discrete space-time predictor solution $\mathbf{q}_h^n$ with its expansion on the nodal basis and after integrating by parts in time, we obtain

$$\begin{aligned} &\int_0^1\int_0^1\int_0^1 \theta_k(\xi,\eta,1)\,\theta_\ell(\xi,\eta,1)\,\hat{\mathbf{q}}_\ell^n d\xi d\eta d\tau \\ &+ \int_0^1\int_0^1\int_0^1 \frac{\partial \theta_k(\xi,\eta,\tau)}{\partial \tau}\theta_\ell(\xi,\eta,\tau)\,\hat{\mathbf{q}}_\ell^n d\xi d\eta d\tau \\ &= \int_0^1\int_0^1\int_0^1 \theta_k(\xi,\eta,0)\,\mathbf{w}_h^n(\xi,\eta,t^n)\,d\xi d\eta d\tau \\ &- \int_0^1\int_0^1\int_0^1 \theta_k \left(\frac{\partial \mathbf{f}^\star(\mathbf{q}_h^n)}{\partial \xi} + \frac{\partial \mathbf{g}^\star(\mathbf{q}_h^n)}{\partial \eta} \right) d\xi d\eta d\tau \\ &+ \int_0^1\int_0^1 \times \int_0^1 \theta_k \left(\mathbf{S}^\star(\mathbf{q}_h^n) - \mathbf{B}_1^\star(\mathbf{q}_h^n) \frac{\partial \mathbf{q}_h^n}{\partial \xi} \right. \\ &\left. - \mathbf{B}_2^\star(\mathbf{q}_h^n) \frac{\partial \mathbf{q}_h^n}{\partial \eta} \right) d\xi d\eta d\tau. \end{aligned} \tag{29}$$

To recover the value of the unknown degrees of freedom $\hat{\mathbf{q}}_\ell^n$, it is sufficient to solve the above equation locally for each element. One important advantage of using the nodal Gauss-Legendre basis is that the terms in (29) can be evaluated in a *dimension-by-dimension* fashion.

2.5.2. Space-Time Predictor for Sliver Space–Time Elements

When a topology change occurs, some space–time sliver elements, as those shown on the right side of **Figure 8**, are originated (see Gaburro et al. [77]), and the predictor procedure over them needs particular care. The problem connected with sliver elements is the fact that their bottom face, which consists only in a line segment, is degenerate, hence the spatial integral over P_i^n vanishes, i.e., there is no possibility to introduce an initial condition for the local Cauchy problem at time t^n into their predictor. Thus, in order to couple however (24) with some known data from the past, we will end up with a formula different from (25). We underline that we *first* carry out the space–time predictor for all standard elements using, which can be computed independently of each other, and only subsequently we process the remaining space–time sliver elements. Then, when

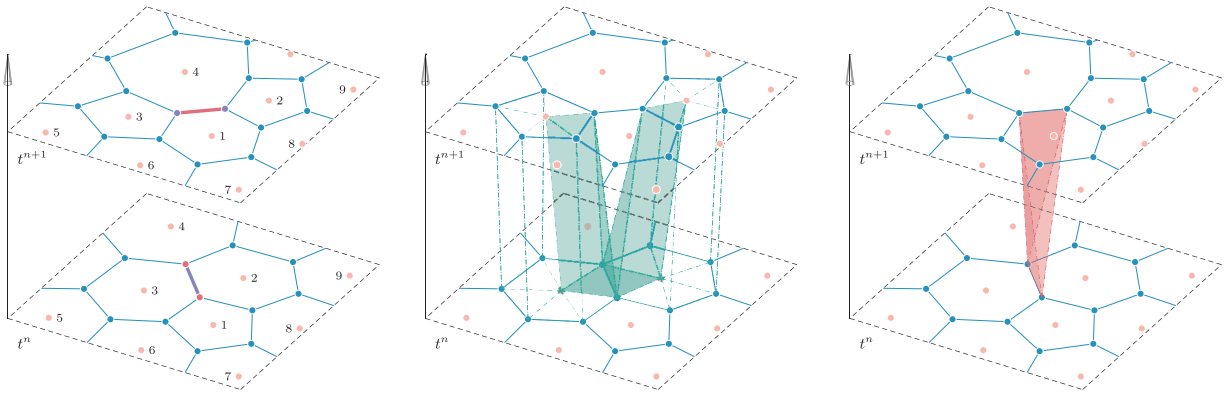


FIGURE 8 | Space time connectivity *with* topology changes and sliver element. Left: at time t^n the polygons P_2^n and P_3^n are neighbors and share the highlighted edge, instead at time t^{n+1} they do not touch each other; the opposite situation occurs for polygons P_1^n and P_4^n. This change of topology causes the appearance of degenerate elements of different types (refer to Gaburro et al. [77] for all the details). In particular, so-called *space–time sliver elements* (right) need to be taken into account when considering the space–time framework, so the predictor and the corrector step have to be a adapted to their special features. Sliver elements (right) are indeed completely new control volumes which do neither exist at time t^n, nor at time t^{n+1}, since they coincide with an edge of the tessellation and, as such, have zero areas in space. However, they have a *non-negligible volume* in space–time. The difficulties associated to this kind of element are due to the fact that $\mathbf{w}_h$ is not clearly defined for it at time t^n (thus the predictor has to be modified) and that contributions across it should not be lost at time t^{n+1} in order to guarantee conservation (thus the corrector has to be modified).

considering a sliver, we use the upwinding in time approach on the entire space–time surface ∂C_i^n that closes a sliver control volume, and again respecting the causality principle, we take the information to feed the predictor only from the past, i.e., only from those space–time neighbors C_j^n whose common surface ∂C_{ij}^n exhibit a *negative* time component of the outward pointing space–time normal vector ($\tilde{\mathbf{n}}_t < 0$). In this way, we can introduce information from the past into the space–time sliver elements.

As a consequence, the predictor solution $\mathbf{q}_h^n$ is again obtained by means of (24), but by treating the *entire* ∂C_i^n with the upwind in time approach, i.e., by considering also the jump terms between the still unknown predictor of the slivers (call it $\mathbf{q}_h^{n,-}$) and the already known predictors of its neighbors (call them $\mathbf{q}_h^{n,+}$),

$$\begin{aligned}
&\int_{C_i^n} \theta_k(\mathbf{x},t)\frac{\partial}{\partial t}\mathbf{q}_h^n(\mathbf{x},t)\,d\mathbf{x}\,dt \\
&- \int_{\partial C_i^-} \theta_k(\mathbf{x},t^n)\left(\mathbf{q}_h^{n,+}-\mathbf{q}_h^{n,-}\right) \\
&- (\mathbf{B}\cdot\tilde{\mathbf{n}})(\mathbf{q}_h^{n,+}-\mathbf{q}_h^{n,-})\big)\,d\mathbf{S}\,dt \\
&+ \int_{C_i^n\setminus\partial C_i^n} \theta_k(\mathbf{x},t)\nabla\cdot\mathbf{F}(\mathbf{q}_h^n)\,d\mathbf{x}\,dt \\
&= \int_{C_i^n\setminus\partial C_i^n} \theta_k(\mathbf{x},t)\left(\mathbf{S}(\mathbf{q}_h^n)-\mathbf{B}(\mathbf{q}_h^n)\cdot\nabla\mathbf{q}_h^n\right)\,d\mathbf{x}\,dt,
\end{aligned} \tag{30}$$

where $\partial C_i^- = \partial C_i^n$ with $\tilde{\mathbf{n}}_t < 0$ is the part of the space-time boundary that has a negative time component of the space-time normal vector. Note that here we have taken into account also the jump of the non-conservative terms, and that these contributions have been added entirely [i.e., not only half of them, as in (49)]. Indeed, in (49) half of the jump contribution goes to one element, while the other half goes to the neighboring element; here instead, since the interaction between neighbors is only computed from the side of the sliver element, the entire jump contributes to the predictor in the sliver element.

2.6. Convergence Proof of the Predictor Step for a Non-linear Conservation Law

In this section, the convergence proof of the predictor for a non-linear conservation law is given. The proof is provided, for simplicity, in the case of a fixed mesh in one space dimension, following the nomenclature already employed in section 2.5.1, but it still holds in higher dimensions. Let us consider a general hyperbolic system of conservation laws of the form

$$\frac{\partial \mathbf{Q}}{\partial t}+\frac{\partial \mathbf{f}}{\partial x}=0. \tag{31}$$

Then, the corresponding space-time DG predictor used in the ADER-DG framework reads

$$\int_0^1\int_0^1 \theta_k\frac{\partial \mathbf{q}_h}{\partial \tau}d\xi\,d\tau+\frac{\Delta t}{\Delta x}\int_0^1\int_0^1 \theta_k\frac{\partial \mathbf{f}_h}{\partial \xi}d\xi\,d\tau=0. \tag{32}$$

For convenience, all derivatives and integrals in (32) have been transformed to the reference space-time element $[0,1]^2$. Moreover, the discrete solution is given by $\mathbf{q}_h=\theta_l(\xi,\tau)\hat{\mathbf{q}}_\ell$, and the flux is expanded in the same basis as $\mathbf{f}_h=\theta_\ell(\xi,\tau)\hat{\mathbf{f}}_\ell$. When using a nodal basis, we can compute the degrees of freedom for the flux interpolant $\mathbf{f}_h$ simply as $\hat{\mathbf{f}}_\ell=\mathbf{f}\left(\hat{\mathbf{q}}_\ell\right)$. We also recall that the initial condition given by the DG scheme at time t^n reads $\mathbf{w}_h=\varphi_\ell(\xi)\hat{\mathbf{w}}_\ell$. Then, integration of the first term in (32) by parts

in time yields

$$\int_0^1 \theta_k(\xi,1)\mathbf{q}_h d\xi - \int_0^1\int_0^1 \frac{\partial \theta_k}{\partial \tau}\mathbf{q}_h d\xi\, d\tau + \frac{\Delta t}{\Delta x}\int_0^1\int_0^1 \theta_k \frac{\partial \mathbf{f}_h}{\partial \xi} d\xi\, d\tau = \int_0^1 \theta_k(\xi,0)\mathbf{w}_h d\xi, \tag{33}$$

and insertion of the definitions of the discrete solution leads to

$$\left(\int_0^1 \theta_k(\xi,1)\theta_l(\xi,1)d\xi - \int_0^1\int_0^1 \frac{\partial \theta_k}{\partial \tau}\theta_l d\xi\, d\tau\right)\hat{\mathbf{q}}_l + \frac{\Delta t}{\Delta x}\int_0^1\int_0^1 \theta_k \frac{\partial \theta_l}{\partial \xi} d\xi\, d\tau\, \hat{\mathbf{f}}_l = \int_0^1 \theta_k(\xi,0)\varphi_l(\xi)d\xi\, \hat{\mathbf{w}}_l. \tag{34}$$

The iterative scheme employed to find the solution for the space-time degrees of freedom $\hat{\mathbf{q}}$, at any Picard iteration r, can therefore be rewritten in compact matrix-vector notation as

$$\mathbf{K}_1\hat{\mathbf{q}}^{r+1} + \frac{\Delta t}{\Delta x}\mathbf{K}_\xi\, \mathbf{f}\left(\hat{\mathbf{q}}^{r+1}\right) = \mathbf{F}_0\hat{\mathbf{w}}^n \tag{35}$$

with

$$\mathbf{K}_1 = \int_0^1 \theta_k(\xi,1)\theta_l(\xi,1)d\xi - \int_0^1\int_0^1 \frac{\partial \theta_k}{\partial \tau}\theta_l d\xi\, d\tau, \tag{36}$$

$$\mathbf{K}_\xi = \int_0^1\int_0^1 \theta_k \frac{\partial \theta_l}{\partial \xi} d\xi\, d\tau, \qquad \mathbf{F}_0 = \int_0^1 \theta_k(\xi,0)\varphi_l(\xi)d\xi, \tag{37}$$

where we have dropped the indices to ease the notation. After inverting $\mathbf{K}_1$ (this matrix is built using the linearly independent basis functions so that it is invertible), we obtain the explicit iteration formula

$$\hat{\mathbf{q}}^{r+1} = \mathbf{K}_1^{-1}\mathbf{F}_0\hat{\mathbf{w}}^n - \frac{\Delta t}{\Delta x}\mathbf{K}_1^{-1}\mathbf{K}_\xi\, \mathbf{f}\left(\hat{\mathbf{q}}^r\right). \tag{38}$$

To prove that the former iterative formula will converge, we introduce the operator

$$\boldsymbol{\varphi}\left(\hat{\mathbf{q}}\right) = \mathbf{K}_1^{-1}\mathbf{F}_0\hat{\mathbf{u}}^n - \frac{\Delta t}{\Delta x}\mathbf{K}_1^{-1}\mathbf{K}_\xi\, \mathbf{f}\left(\hat{\mathbf{q}}\right), \tag{39}$$

and the induced matrix norm

$$\|\mathbf{A}\| = \sup_{\mathbf{x}\neq 0}\frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|}. \tag{40}$$

Furthermore, we assume the flux to be Lipschitz continuous with Lipschitz constant $L > 0$ so that

$$\left\|\mathbf{f}\left(\hat{\mathbf{p}}\right) - \mathbf{f}\left(\hat{\mathbf{q}}\right)\right\| \leq L\left\|\hat{\mathbf{p}} - \hat{\mathbf{q}}\right\|. \tag{41}$$

We now need to show that the operator $\boldsymbol{\varphi}$ is a contraction:

$$\begin{aligned}\left\|\boldsymbol{\varphi}\left(\hat{\mathbf{q}}\right) - \boldsymbol{\varphi}\left(\hat{\mathbf{p}}\right)\right\| &= \left\|\mathbf{K}_1^{-1}\mathbf{F}_0\hat{\mathbf{u}}^n - \mathbf{K}_1^{-1}\mathbf{F}_0\hat{\mathbf{u}}^n - \frac{\Delta t}{\Delta x}\mathbf{K}_1^{-1}\mathbf{K}_\xi\, \mathbf{f}\left(\hat{\mathbf{q}}\right) + \frac{\Delta t}{\Delta x}\mathbf{K}_1^{-1}\mathbf{K}_\xi\, \mathbf{f}\left(\hat{\mathbf{p}}\right)\right\| \\ &= \frac{\Delta t}{\Delta x}\left\|\mathbf{K}_1^{-1}\mathbf{K}_\xi\, \left(\mathbf{f}\left(\hat{\mathbf{p}}\right) - \mathbf{f}\left(\hat{\mathbf{q}}\right)\right)\right\| \\ &\leq \frac{\Delta t}{\Delta x}\left\|\mathbf{K}_1^{-1}\mathbf{K}_\xi\right\|\, \left\|\mathbf{f}\left(\hat{\mathbf{p}}\right) - \mathbf{f}\left(\hat{\mathbf{q}}\right)\right\| \\ &\leq L\frac{\Delta t}{\Delta x}\left\|\mathbf{K}_1^{-1}\mathbf{K}_\xi\right\|\, \left\|\hat{\mathbf{p}} - \hat{\mathbf{q}}\right\|. \end{aligned} \tag{42}$$

The operator is therefore a *contraction* under the CFL-type condition on the time step Δt

$$0 < L\frac{\Delta t}{\Delta x}\left\|\mathbf{K}_1^{-1}\mathbf{K}_\xi\right\| < 1, \tag{43}$$

which connects the Lipschitz constant L with the mesh spacing Δx and the matrix norm of $\left\|\mathbf{K}_1^{-1}\mathbf{K}_\xi\right\|$. Since the operator is contractive under the above assumptions, the Banach fixed point theorem, Banach [144], guarantees convergence of the iterative method.

In the previous reasoning, we have assumed that the inequality in the right hand side of (43) be strict. Thus, to conclude the proof, let us assume that the equality holds, this is true if and only if $\left\|\mathbf{K}_1^{-1}\mathbf{K}_\xi\right\| = 0$. By taking into account the definition of the induced matrix norm (40), it implies $\left\|\mathbf{K}_1^{-1}\mathbf{K}_\xi\, \mathbf{x}\right\| = 0$ for any $\mathbf{x}$ in the metric space. Thus, $\mathbf{K}_1^{-1}\mathbf{K}_\xi = 0$. Direct substitution in (38) gives

$$\mathbf{K}_1\hat{\mathbf{q}}^{r+1} = \mathbf{F}_0\hat{\mathbf{w}}^n, \tag{44}$$

so that no iterative procedure is done.

Note: The matrix $\mathbf{K}_1^{-1}\mathbf{K}_\xi$ has been proven to be nilpotent and thus all its eigenvalues are zero, see Jackson [143], which guarantees convergence to the exact solution in a finite number of steps for linear homogeneous PDE.

2.7. Corrector Step

The corrector step is the last step of our *path-conservative* ADER FV-DG scheme, where the update of the solution from time t^n up to time t^{n+1} can take place in a single step procedure thanks to the use of the predictor $\mathbf{q}_h^n$.

The update formula is recovered starting from the space–time divergence form of the PDE

$$\tilde{\nabla}\cdot\tilde{\mathbf{F}}(\mathbf{Q}) + \tilde{\mathbf{B}}(\mathbf{Q})\cdot\tilde{\nabla}\mathbf{Q} = \mathbf{S}(\mathbf{Q}), \quad \tilde{\mathbf{F}} = (\mathbf{F},\mathbf{Q}), \quad \tilde{\mathbf{B}} = (\mathbf{B},\mathbf{0}), \text{ and } \tilde{\nabla} = (\partial_{\mathbf{x}},\, \partial_t)^T, \tag{45}$$

which is multiplied by a set of space–time test functions $\tilde{\varphi}_k$ and integrated over each space–time control volume C_i^n

$$\int_{C_i^n}\tilde{\varphi}_k(\mathbf{x},t)\left(\nabla\cdot\tilde{\mathbf{F}}(\mathbf{Q}) + \tilde{\mathbf{B}}(\mathbf{Q})\cdot\nabla\mathbf{Q}\right)\, d\mathbf{x}\, dt = \int_{C_i^n}\tilde{\varphi}_k(\mathbf{x},t)\mathbf{S}(\mathbf{Q})\, d\mathbf{x}\, dt. \tag{46}$$

Note that the employed test functions $\tilde{\varphi}_k$ coincide with the θ_k of (22) for the Cartesian Case A. Instead, for the moving polygonal Case B, they need to be tied to the motion of the barycenter $\mathbf{x}_{\mathbf{b}_i}(t)$ and must be moved together with $P_i(t)$ in such a way that at time $t = t^n$ they refer to the current barycenter $\mathbf{x}_{b_i}^n$ and at time $t = t^{n+1}$ they refer to the new barycenter $\mathbf{x}_{b_i}^{n+1}$, thus they are defined as follows

$$\begin{aligned} &\tilde{\varphi}_\ell(x,y,t)|_{C_i^n} = \frac{(x-x_{b_i}(t))^{p_\ell}}{p_\ell!\, h_i^{p_\ell}} \frac{(y-y_{b_i}(t))^{q_\ell}}{q_\ell!\, h_i^{q_\ell}}, \\ &\text{with } \mathbf{x}_{b_i}(t) = \frac{t-t^n}{\Delta t}\mathbf{x}_{b_i}^n + \left(1-\frac{t-t^n}{\Delta t}\right)\mathbf{x}_{b_i}^{n+1}, \\ &\ell = 0,\ldots,\mathcal{N}, \quad 0 \le p+q \le N. \end{aligned} \tag{47}$$

These moving modal basis functions are essential to the moving approach presented in Gaburro et al. [77] and used in this paper. They naturally allow for topology changes, without the need of any remapping steps, which we want to avoid in a direct ALE formulation.

Now, (46) by applying the Gauss theorem to the flux-divergence term and by splitting the non-conservative products into their volume and surface contribution, becomes

$$\begin{aligned} &\int_{P_i^{n+1}} \tilde{\varphi}_k \mathbf{u}_h(\mathbf{x},t^{n+1})\, d\mathbf{x} = \int_{P_i^n} \tilde{\varphi}_k \mathbf{u}_h(\mathbf{x},t^n)\, d\mathbf{x} \\ &- \sum_{j=1}^{N_{V_i}^{n,st}} \int_{\partial C_{ij}^n} \tilde{\varphi}_k \mathcal{D}(\mathbf{q}_h^{n,-},\mathbf{q}_h^{n,+})\cdot\tilde{\mathbf{n}}\, dS \\ &+ \int_{C_i^n\setminus\partial C_i^n} \tilde{\nabla}\tilde{\varphi}_k \cdot \tilde{\mathbf{F}}(\mathbf{q}_h)\, d\mathbf{x}dt \\ &+ \int_{C_i^n\setminus\partial C_i^n} \tilde{\varphi}_k(\mathbf{x},t)\left(\mathbf{S}(\mathbf{q}_h^n) - \mathbf{B}(\mathbf{q}_h^n)\cdot\nabla\mathbf{q}_h^n\right) d\mathbf{x}\, dt, \end{aligned} \tag{48}$$

where $\mathbf{Q}$ on P_i^{n+1} is represented by the unknown $\mathbf{u}_h^{n+1}$, on P_i^n is taken to be the current representation of the conserved variables $\mathbf{u}_h^n$, in the interior of C_i^n is given by the predictor $\mathbf{q}_h^n$ and on the space–time lateral surfaces ∂C_{ij}^n is given by $\mathbf{q}_h^{n,-}$ and $\mathbf{q}_h^{n,+}$ which are the so-called boundary-extrapolated data, i.e., the values assumed respectively by the predictors of the two neighbor elements C_i^n and C_j^n on the shared space–time lateral surface ∂C_{ij}^n. Furthermore, we have employed a two-point path-conservative numerical flux function of Rusanov-type

$$\begin{aligned} \mathcal{D}(\mathbf{q}_h^{n,-},\mathbf{q}_h^{n,+})\cdot\tilde{\mathbf{n}} =\ & \frac{1}{2}\left(\tilde{F}(\mathbf{q}_h^{n,+}) + \tilde{\mathbf{F}}(\mathbf{q}_h^{n,-})\right)\cdot\tilde{\mathbf{n}} \\ -\frac{1}{2}s_{\max}&\left(\mathbf{q}_h^{n,+}-\mathbf{q}_h^{n,-}\right) \\ +\frac{1}{2}&\left(\int_0^1 \tilde{B}\left(\Psi(\mathbf{q}_h^{n,-},\mathbf{q}_h^{n,+},s)\right)\cdot\mathbf{n}\, d\mathbf{x}\right)\cdot\left(\mathbf{q}_h^{n,+}-\mathbf{q}_h^{n,-}\right), \end{aligned} \tag{49}$$

where $s_{\max}$ is the maximum eigenvalue of the ALE Jacobian matrices $\mathbf{A}_{\mathbf{n}}^{\mathbf{V}}(\mathbf{q}_h^{n,+})$ and $\mathbf{A}_{\mathbf{n}}^{\mathbf{V}}(\mathbf{q}_h^{n,-})$ being

$$\mathbf{A}_{\mathbf{n}}^{\mathbf{V}}(\mathbf{Q}) = \left(\sqrt{\tilde{n}_x^2+\tilde{n}_y^2}\right)\left[\frac{\partial\mathbf{F}}{\partial\mathbf{Q}}\cdot\mathbf{n} - (\mathbf{V}\cdot\mathbf{n})\,\mathbf{I}\right], \mathbf{n} = \frac{(\tilde{n}_x,\tilde{n}_y)^T}{\sqrt{\tilde{n}_x^2+\tilde{n}_y^2}}, \tag{50}$$

and the path $\Psi = \Psi(\mathbf{q}_h^-,\mathbf{q}_h^+,s)$ is a straight-line segment path

$$\psi = \psi(\mathbf{q}_h^-,\mathbf{q}_h^+,s) = \mathbf{q}_h^- + s\left(\mathbf{q}_h^+ - \mathbf{q}_h^-\right), s\in[0,1], \tag{51}$$

connecting $\mathbf{q}_h^{n,-}$ and $\mathbf{q}_h^{n,+}$ which allow to treat the jump of the non-conservative products following the theory introduced in Dal Maso et al. [145], Parés [146], and Castro et al. [147], and extended to ADER FV-DG schemes of arbitrary high order in Dumbser et al. [46] and Dumbser and Toro [148]. Despite in this paper we only consider the Rusanov flux, the above methodology can be extended to different flux functions, adapting to the new flux splitting techniques like the ones presented in Toro and Vázquez-Cendón [149]. Finally, the time step size Δt is given by

$$\begin{aligned} \Delta t &< \mathrm{CFL}\frac{h_{\min}}{(2N+1)}\frac{1}{|\lambda_{\max}|}, \quad \text{(Case A)}, \\ \Delta t &< \mathrm{CFL}\left(\frac{|P_i^n|}{(2N+1)\,|\lambda_{\max}|\,\sum_{\partial P_{ij}^n}|\ell_{ij}|}\right) \quad \text{(Case B)}, \end{aligned} \tag{52}$$

where $h_{\min}$ is the minimum characteristic mesh-size, ℓ_{ij} is the length of the edge j of P_i^n and $|\lambda_{\max}|$ is the spectral radius of the Jacobian of the flux $\mathbf{F}$. Stability on unstructured meshes is guaranteed by the satisfaction of the inequality $\mathrm{CFL} < \frac{1}{d}$, see Dumbser et al. [42].

We close this section by remarking that the integration of the governing PDE over *closed* space-time volumes C_i^n automatically satisfies the geometric conservation law (GCL) for all test functions $\tilde{\varphi}_k$. This simply follows from the Gauss theorem and we refer to Boscheri and Dumbser [63] for a complete proof.

2.8. A Posteriori Subcell Finite Volume Limiter

Up to now, we have presented a family of FV and DG type schemes which achieves arbitrary high order of accuracy in space and time; the main difference between the FV and the DG approach lies in the fact that FV schemes, thanks to the WENO-type non-linear reconstruction procedure, are robust in the presence of shocks and discontinuities, while the DG formulation as presented so far, being linear in the sense of Godunov, is subject to the appearance of spurious oscillations. Thus, in order to employ a DG scheme in the context of solving hyperbolic partial differential equations, where usually discontinuities are developed, a technique that is able to limit spurious oscillations (called *limiter*) should be introduced. Several attempts in that direction can be found in the literature. For example, we could recall the *artificial viscosity* technique used in Hartmann and Houston [150],Persson and Peraire [151], and Cesenek et al. [152] which consists in adding a small parabolic term in the equation in order to smooth out the discontinuities.

Here, instead, we follow a different approach based on exploiting the respective strengths of FV and DG schemes, i.e., the resolution of DG in smooth regions and the robustness of FV across discontinuities. Thus, we first evolve the solution everywhere by using our DG scheme; then, we check *a posteriori*, at the end of each time step, if the obtained DG solution in each cell respects or not some criteria [as density and pressure positivity, a relaxed discrete maximum principle, specific physical bounds, or more elaborate choices as those of Guermond et al. [153]], and we mark as *troubled* those cells where the obtained DG solution is marked as not acceptable. Only for these troubled cells we repeat the time step using, instead of the DG scheme, a second order TVD FV method, which always assures a robust solution.

This idea is founded on works as those of Cockburn and Shu [154], Qiu and Shu [155, 156], Balsara et al. [157], Luo et al. [158], Krivodonova [159], Zhu et al. [160], Zhu and Qiu [161], Clain et al. [86], Diot et al. [87, 88], Loubére et al. [79], Boscheri et al. [162],and Boscheri and Loubére [83]; but in particular, here, we adopt a so-called *subcell* approach aimed at not losing the resolution of the DG scheme when switching to the FV method, as forwarded in Sonntag and Munz [163], Dumbser et al. [78], Zanotti et al. [80], Dumbser and Loubére [81], Boscheriand Dumbser [119], Fambri et al. [84], Rannabauer et al. [164], de la Rosa and Munz [165], and Boscheri et al. [142]. Indeed, at the beginning of the time step we *project* the DG solution $\mathbf{u}_h^n$ of a troubled cell P_i^n on a subdivision of it in sub-cells $s_{i,\alpha}^n$ obtaining a value for the cell averages on $s_{i,\alpha}^n$ at time t^n

$$\mathbf{v}_{i,\alpha}^n(\mathbf{x},t^n) = \frac{1}{|s_{i,\alpha}^n|}\int_{s_{i,\alpha}^n} \mathbf{u}_h^n(\mathbf{x},t^n)\,d\mathbf{x} = \frac{1}{|s_{i,\alpha}^n|}\int_{s_{i,\alpha}^n} \varphi_\ell(\mathbf{x})\,d\mathbf{x}\,\hat{\mathbf{u}}_l^n = \mathcal{P}(\mathbf{u}_h^n) \qquad \forall\alpha. \tag{53}$$

We evolve the cell averages up to time t^{n+1} using a classical TVD FV scheme, obtaining $\mathbf{v}_{i,\alpha}^{n+1}(\mathbf{x},t^{n+1})$. Finally, we recover a DG polynomial representation of the solution at time t^{n+1} over P_i^{n+1} using the values on the sub-grid level $\mathbf{v}_{i,\alpha}^{n+1}$ and by applying a *reconstruction* operator as

$$\int_{S_{i,\alpha}^n} \mathbf{u}_h^{n+1}(\mathbf{x},t^{n+1})\,d\mathbf{x} = \int_{S_{i,\alpha}^n} \mathbf{v}_{i,\alpha}^{n+1}(\mathbf{x},t^n)\,d\mathbf{x} = \mathcal{R}(\mathbf{v}_{i,\alpha}^{n+1}(\mathbf{x},t^n))\forall\alpha, \tag{54}$$

where the reconstruction is imposed to be *conservative* on the main cell P_i^{n+1} yielding the additional linear constraint

$$\int_{P_i^{n+1}} \mathbf{u}_h(\mathbf{x},t^{n+1})\,d\mathbf{x} = \int_{P_i^{n+1}} \mathbf{v}_h(\mathbf{x},t^{n+1})\,d\mathbf{x}. \tag{55}$$

Thus, the limited solution on a troubled cell is *robust* thanks to the use of a TVD scheme and *accurate* thanks to the subcell resolution.

For all the details of the *a posteriori* subcell FV limiter used in this work, we refer to Dumbser et al. [78] and Fambri et al. [36] for the fixed Cartesian Case A and to Gaburro et al. [77] for the moving polygonal Case B.

3. A UNIFIED FIRST ORDER HYPERBOLIC MODEL OF CONTINUUM MECHANICS

3.1. Governing PDE System

A simplified diffuse interface formulation of the unified continuum fluid and solid mechanics model [57, 59, 60, 166], which can be used for modeling moving boundary problems of fluids and solids of arbitrary geometry, is given by the following PDE system (throughout this paper we make use of the Einstein summation convention over repeated indices)

$$\frac{\partial \alpha}{\partial t} + v_k\frac{\partial \alpha}{\partial x_k} = 0, \tag{56a}$$

$$\frac{\partial (\alpha\rho)}{\partial t} + \frac{\partial (\alpha\rho v_k)}{\partial x_k} = 0, \tag{56b}$$

$$\frac{\partial (\alpha\rho v_i)}{\partial t} + \frac{\partial \left(\alpha\rho v_i v_k + \alpha p\delta_{ik} - \alpha\sigma_{ik}\right)}{\partial x_k} = \rho g_i, \tag{56c}$$

$$\frac{\partial A_{ik}}{\partial t} + \frac{\partial (A_{ij}v_j)}{\partial x_k} + v_j\left(\frac{\partial A_{ik}}{\partial x_j} - \frac{\partial A_{ij}}{\partial x_k}\right) = -\frac{1}{\theta_1(\tau_1)}E_{A_{ik}}, \tag{56d}$$

$$\frac{\partial (\alpha\rho J_i)}{\partial t} + \frac{\partial \left(\alpha\rho J_i v_k + T\delta_{ik}\right)}{\partial x_k} = -\frac{1}{\theta_2(\tau_2)}E_{J_i}, \tag{56e}$$

$$\frac{\partial (\alpha\rho S)}{\partial t} + \frac{\partial \left(\alpha\rho S v_k + E_{J_k}\right)}{\partial x_k} = \frac{\rho}{T}\left(\frac{1}{\theta_1}E_{A_{ik}}E_{A_{ik}} + \frac{1}{\theta_2}E_{J_k}E_{J_k}\right) \geq 0, \tag{56f}$$

$$\frac{\partial (\alpha\rho E)}{\partial t} + \frac{\partial \left(v_k\alpha\rho E + \alpha v_i(p\delta_{ik} - \sigma_{ik})\right)}{\partial x_k} = \rho g_i v_i. \tag{56g}$$

Here, (56a) is the evolution equation for the color function α that is needed in the diffuse interface approach as introduced in Tavelli et al. [85] for the description of linear elastic solids of arbitrary geometry and as used in Dumbser [106] and Gaburro et al. [107] for a simple diffuse interface method for the simulation of non-hydrostatic free surface flows. We assume that the color function α equals to 1 in the regions of the computational domain occupied by the material and 0 outside these regions. In the computational code, $\alpha = 1-\varepsilon$ inside of the material and $\alpha = \varepsilon$ outside the material. Here, ε is a small parameter $\varepsilon \ll 1$, see section 4. Then, inside of the diffuse interface, α may take any values between 0 and 1 (between ε and $1-\varepsilon$ in the computational code). Equation (56b) is the mass conservation law and ρ is the material density; (56c) is the momentum conservation law, where v_i is the velocity field and g_i is the gravity vector; (56d) is the evolution equation for distortion field A_{ik} (non-holonomic basis triad, see Peshkov et al. [167]); (56e) is the evolution equation for the specific thermal impulse J_k constituting the heat conduction in the matter via a hyperbolic (non-Fourier–type) model. Finally, (56f) is the entropy balance equation and (56g) is the energy conservation law. Other thermodynamic parameters are defined via the total energy potential $E = E(\alpha,\rho,S,\boldsymbol{v},\boldsymbol{A},\boldsymbol{J})$: $\Sigma_{ik} = p\delta_{ik} - \sigma_{ik}$ is the total stress tensor (δ_{ik} is the Kronecker delta); $p = \rho^2 E_\rho$ is the thermodynamic pressure; $\sigma_{ik} = -\rho A_{jk}E_{A_{ji}}$ is the non-isotropic part of the stress tensor, $T = E_S$ is the temperature, and the

notations such as E_ρ, $E_{A_{ik}}$, etc. stand for the partial derivatives of the energy potential, e.g., $E_\rho = \frac{\partial E}{\partial \rho}$, $E_{A_{ik}} = \frac{\partial E}{\partial A_{ik}}$, etc.

The dissipation in the medium includes two relaxation processes: the shear stress relaxation characterized by the scalar function $\theta_1(\tau_1) > 0$ depending on the relaxation time τ_1 and thermal impulse relaxation characterized by $\theta_2(\tau_2) > 0$ depending on the relaxation time τ_2. Both these relaxation processes then contribute to the entropy production term [the source on the right hand-side of (56f)] which is positive because it is quadratic in $E_{A_{ik}}$ and E_{J_k}.

From the mathematical standpoint, the unification of the model (56) consists in the use of only first-order hyperbolic equations for both dissipative and non-dissipative processes in contrast to the classical continuum mechanics relying on the mixed hyperbolic-parabolic formulations such as the famous Navier-Stokes-Fourier equations, for example. From the physical standpoint, the unification of Equations (56) consists in treating solid and fluid states of matter from the solid-dynamics viewpoint. Indeed, as discussed in Peshkov and Romenski [57] and Dumbser et al. [59, 166], similarly to standard continuum solid-dynamics, the distortion field introduces additional degrees of freedom (in comparison to the classical continuum fluid mechanics) which characterizes deformation and rotational degrees of freedom of the continuum particles, represented not as scaleless mathematical points but characterized by a finite length scale, or equivalently, time scale τ_1, e.g., see Dumbser et al. [166]. In such a formulation, solid-type behavior corresponds to relaxation times τ_1 such that $T^{problem} \ll \tau_1$, while the fluid-type behavior corresponds to $\tau_1 \ll T^{problem}$, where $T^{problem}$ is the characteristic time scale of the problem under consideration.

In order to close system (56), that is, in order to define pressure $p = \rho^2 E_\rho$, stresses $\sigma_{ik} = -\rho A_{jk} E_{A_{ji}}$, temperature $T = E_S$, and the dissipative source terms, one needs to provide the energy potential E. In this paper, we rely on a rather simple choice of E, which is, however, enough to deal with Newtonian fluids and simple hyperelastic solids. Thus, we assume that the specific total energy can be written as a sum of three contributions as

$$E(\alpha, \rho, S, v_i, A_{ik}, J_k) = E_1(\rho, S) + E_2(\alpha, A_{ik}, J_k) + E_3(v_i), \quad (57)$$

with the specific internal energy given by the ideal gas equation of state

$$E_1(\rho, S) = \frac{c_0^2}{\gamma(\gamma-1)},\ c_0^2 = \gamma \rho^{\gamma-1} e^{S/c_v}, \text{ or}$$
$$E_1(\rho, p) = \frac{p}{\rho(\gamma-1)}, \quad (58)$$

in the case of gases, and given by either the so-called stiffened gas equation of state

$$E_1(\rho, S) = \frac{c_0^2}{\gamma(\gamma-1)} \left(\frac{\rho}{\rho_0}\right)^{\gamma-1} e^{S/c_v} + \frac{\rho_0 c_0^2 - \gamma p_0}{\gamma \rho} \quad (59)$$

or the well-known Mie-Grüneisen equation of state

$$E_1(\rho, p) = \frac{p - \rho_0 c_0^2 f(\nu)}{\rho_0 \Gamma_0},$$
$$f(\nu) = \frac{(\nu-1)(\nu - \frac{1}{2}\Gamma_0(\nu-1))}{(\nu - s(\nu-1))^2}, \quad \nu = \frac{\rho}{\rho_0}, \quad (60)$$

in the case of solids and liquids. Here, c_v is the specific heat capacity at constant volume, γ is the ratio of the specific heats, p_0 is the reference (atmospheric) pressure, ρ_0 is the reference material density, and Γ_0, and s are some material parameters. The specific energy stored in material deformations and in the thermal impulse is

$$E_2(\alpha, A_{ik}, J_k) = \frac{1}{4}\bar{c}_s^2\, \mathring{G}_{ij}\mathring{G}_{ij} + \frac{1}{2}\bar{c}_h^2 J_k J_k, \quad (61)$$

where $\mathring{G}_{ij} = G_{ij} - \frac{1}{3}G_{kk}\delta_{ij}$ is the trace-free part of the metric tensor $G_{ij} = A_{ki}A_{kj}$, which is induced by the mapping from Eulerian coordinates to the current stress-free reference configuration. The coefficients $\bar{c}_s(\alpha)$ and $\bar{c}_h(\alpha)$ in (61) are the characteristic velocities for propagation of shear and thermal perturbations accordingly. In the present diffuse interface model, we choose the following simple linear mixture rule for the computation of the shear sound speed and of the heat wave propagation as a function of the volume fraction α

$$\bar{c}_s(\alpha) = \alpha c_s + (1-\alpha)c_s^g, \qquad \bar{c}_h(\alpha) = \alpha c_h + (1-\alpha)c_h^g, \quad (62)$$

where c_s and c_h are the material parameters inside the continuum and $c_h^g \ll 1$ and $c_s^g \ll 1$ are free parameters that can be chosen for the region outside the continuum. The specific kinetic energy is contained in the third contribution to the total energy and reads $E_3(v_k) = \frac{1}{2}v_i v_i$.

With the equation of state chosen above, we get the following expressions for the stress tensor, the heat flux and the dissipative sources $E_{A_{ik}}$ and E_{J_k} present in the relaxation source terms:

$$\sigma_{ik} = \rho \bar{c}_s^2 G_{ij}\mathring{G}_{jk}, \qquad q_k = \rho T \bar{c}_h^2 J_k, \quad (63)$$

$$E_{A_{ik}} = \bar{c}_s^2 A_{ij}\mathring{G}_{jk}, \qquad E_{J_k} = \bar{c}_h^2 J_k. \quad (64)$$

The functions θ_1 and θ_2 are chosen in such a way that a *constant* viscosity and heat conduction coefficient are obtained in the stiff relaxation limit, see Dumbser et al. [59] for a formal asymptotic analysis,

$$\theta_1(\tau_1) = \frac{1}{3}\tau_1 \bar{c}_s^2 |\boldsymbol{A}|^{\frac{5}{3}}, \qquad \theta_2(\tau_2) = \tau_2 \bar{c}_h^2 \frac{\rho}{\rho_0}\frac{T_0}{T}. \quad (65)$$

Thus, following the procedure detailed in Dumbser et al. [59], one can show via formal asymptotic expansion that in the stiff relaxation limit $\tau_1 \to 0$, $\tau_2 \to 0$, the stress tensor and the heat flux reduce to

$$\boldsymbol{\sigma} = -\frac{1}{6}\rho_0 \bar{c}_s^2 \tau_1 \left(\nabla \boldsymbol{v} + \nabla \boldsymbol{v}^T - \frac{2}{3}(\nabla \cdot \boldsymbol{v})\boldsymbol{I}\right) \quad (66)$$

and

$$\boldsymbol{q} = -\bar{c}_h^2 \tau_2 \frac{T_0}{\rho_0} \nabla T, \tag{67}$$

that is the effective shear viscosity and effective heat conductivity of model (56) are

$$\mu = \frac{1}{6} \rho_0 \tau_1 \bar{c}_h^2, \qquad \kappa = \tau_2 \bar{c}_h^2 \frac{T_0}{\rho_0} \tag{68}$$

with ρ_0 and T_0 are reference density and temperature, see Dumbser et al. [59], where also an explanation has been provided of how the relaxation times τ could be obtained experimentally via ultrasound measurements.

3.2. Symmetric Godunov Form of the Model

It is important to note an interesting structural feature of Equations (56) that may affect future developments of the ADER schemes in an attempt to respect such structural properties at the discrete level that may help to improve physical consistency of the numerical solution. Thus, as many PDE systems studied in some other of our papers [59, 60, 168, 169], system (56) belongs to the class of so-called Symmetric Hyperbolic Thermodynamically Compatible (SHTC) PDE systems originally studied by Godunov [170, 171] and later by Godunov and Romenski [172], Godunov et al. [173], Romenski [168] and Romensky [174].

Indeed, by simply rescaling the quantities $\bar{\rho} = \alpha\rho$, $\bar{p} = \alpha p = \bar{\rho}^2 E_{\bar{\rho}}$, and $\bar{\sigma}_{ik} = \alpha\sigma_{ik} = -\bar{\rho} A_{jk} E_{A_{ji}}$ and replacing the non-conservative Equation (56a) by an equivalent (on smooth solutions) conservative form (69a), system (56) can be written as

$$\frac{\partial(\alpha\bar{\rho})}{\partial t} + \frac{\partial(\alpha\bar{\rho}v_k)}{\partial x_k} = 0, \tag{69a}$$

$$\frac{\partial\bar{\rho}}{\partial t} + \frac{\partial(\bar{\rho}v_k)}{\partial x_k} = 0, \tag{69b}$$

$$\frac{\partial(\bar{\rho}v_i)}{\partial t} + \frac{\partial\left(\bar{\rho}v_i v_k + \bar{p}\delta_{ik} - \bar{\sigma}_{ik}\right)}{\partial x_k} = 0, \tag{69c}$$

$$\frac{\partial A_{ik}}{\partial t} + \frac{\partial(A_{ij}v_j)}{\partial x_k} + v_j\left(\frac{\partial A_{ik}}{\partial x_j} - \frac{\partial A_{ij}}{\partial x_k}\right) = -\frac{1}{\theta_1} E_{A_{ik}}, \tag{69d}$$

$$\frac{\partial(\bar{\rho}J_i)}{\partial t} + \frac{\partial\left(\bar{\rho}J_i v_k + E_S\delta_{ik}\right)}{\partial x_k} = -\frac{1}{\theta_2} E_{J_i}, \tag{69e}$$

$$\frac{\partial(\bar{\rho}S)}{\partial t} + \frac{\partial\left(\bar{\rho}S v_k + E_{J_k}\right)}{\partial x_k} = \frac{\bar{\rho}}{\alpha T}\left(\frac{1}{\theta_1} E_{A_{ik}} E_{A_{ik}} + \frac{1}{\theta_2} E_{J_k} E_{J_k}\right) \geq 0, \tag{69f}$$

where we have omitted the energy equation. Now, this system looks exactly as the system studied in Dumbser et al. [59], apart from the additional Equation (69a) which has the same structure as (69b) and does not change the essence. Then, after denoting $\mathcal{E} = \bar{\rho}E$ and introducing new variables $\mathbf{P} = (\varrho_1, \varrho_2, v_i, \alpha_{ik}, \Theta_i, \sigma)$

$$\varrho_1 = \mathcal{E}_{\alpha\bar{\rho}},\ \varrho_2 = \mathcal{E}_{\bar{\rho}},\ v_i = \mathcal{E}_{\bar{\rho}v_i},\ \alpha_{ik} = \mathcal{E}_{A_{ik}},\ \Theta_i = \mathcal{E}_{\bar{\rho}J_i},\ T = \mathcal{E}_{\bar{\rho}S}, \tag{70}$$

which are thermodynamically conjugate to the conservative variables $\mathbf{Q} = (\alpha\bar{\rho}, \bar{\rho}, \bar{\rho}v_i, A_{ik}, \bar{\rho}J_i, \bar{\rho}S)$, and a new thermodynamic potential $L(\mathbf{P}) = \mathbf{Q}\cdot\mathcal{E}_{\mathbf{Q}} - \mathcal{E} = \mathbf{Q}\cdot\mathbf{P} - \mathcal{E}$, system (69) can be written in a symmetric form

$$\frac{\partial L_{\varrho_i}}{\partial t} + \frac{\partial(v_k L)_{\varrho_i}}{\partial x_k} = 0, \quad i = 1, 2, \tag{71a}$$

$$\frac{\partial L_{v_i}}{\partial t} + \frac{\partial(v_k L)_{v_i}}{\partial x_k} + L_{\alpha_{ij}}\frac{\partial\alpha_{kj}}{\partial x_k} - L_{\alpha_{jk}}\frac{\partial\alpha_{jk}}{\partial x_i} = \rho g_i, \tag{71b}$$

$$\frac{\partial L_{\alpha_{il}}}{\partial t} + \frac{\partial(v_k L)_{\alpha_{il}}}{\partial x_k} + L_{\alpha_{jl}}\frac{\partial v_j}{\partial x_i} - L_{\alpha_{il}}\frac{\partial v_k}{\partial x_k} = -\frac{1}{\theta_1}\alpha_{il}, \tag{71c}$$

$$\frac{\partial L_{\Theta_i}}{\partial t} + \frac{\partial(v_k L)_{\Theta_i}}{\partial x_k} + \frac{\partial T}{\partial x_i} = -\frac{1}{\theta_2}\Theta_i, \tag{71d}$$

$$\frac{\partial L_T}{\partial t} + \frac{\partial(v_k L)_T}{\partial x_k} + \frac{\partial\Theta_k}{\partial x_k} = \frac{\varrho_2^2}{\varrho_1 T}\left(\frac{1}{\theta_1}\alpha_{ik}\alpha_{ik} + \frac{1}{\theta_2}\Theta_k\Theta_k\right) \geq 0. \tag{71e}$$

In this PDE system, the first two terms in each equation form the canonical Godunov form introduced in Godunov [170] which can be immediately written as a quasilinear symmetric form, e.g., see Peshkov et al. [169], Romenski [168], and Romensky [174]. The other (non-conservative) terms obviously form a symmetric matrix. Therefore, the entire system (71) can be written in a symmetric quasi-linear form and hence, it is a symmetric hyperbolic system if the thermodynamic potential L is convex.

We note that the understanding of the structural properties of the continuous equations might be beneficial for developing of so-called structure-preserving numerical integrators (e.g., symplectic integrators). Thus, the energy conservation law (56g) is in fact a consequence of the other Equations (56) or (71), e.g., see Dumbser et al. [59] and Peshkov et al. [169], and can be viewed as a constraint of the system (71). Its non-violation at the discrete level cannot be guaranteed by the general purpose ADER family of schemes studied in this paper and hence, usually, as well as in our implementation, it is included into the set of discretized PDEs instead of the entropy equation. In principle, a structure-preserving scheme which satisfies all SHTC properties [169] of the continuous equations at the discrete level should guarantee the automatic satisfaction of the energy conservation law, without its explicit discretization. We hope to cover this topic in future work.

4. NUMERICAL RESULTS

In this section, we present some numerical results in order to illustrate the capabilities and potential applicability of the proposed numerical approach in non-linear continuum mechanics. The first three test problems are carried out without making explicit use of the diffuse interface approach, i.e., setting $\alpha = 1$ everywhere in the entire computational domain. The last three test problems illustrate the full potential of the diffuse interface extension of the GPR model in the context of moving free boundary problems. Gravity effects are neglected in all test cases, apart from the dambreak problem shown in subsection 4.6. Whenever values for $\nu = \mu/\rho_0$ and c_s are provided, the corresponding relaxation time τ_1 is computed according to (68).

4.1. Numerical Convergence Studies in the Stiff Relaxation Limit

In order to verify the high order property of our ADER schemes in both space and time in the stiff relaxation limit, we first represent the numerical convergence study that was already carried out in Dumbser et al. [59] on a smooth unsteady flow, for which an exact analytical solution is known for the compressible Euler equations, i.e., in the stiff relaxation limit $\tau_1 \rightarrow 0$ and $\tau_2 \rightarrow 0$ of the GPR model. The problem setup is the one of the classical isentropic vortex, see Hu and Shu [175]. The initial condition consists in a stationary isentropic vortex, whose exact solution can easily be found by solving the compressible Euler equations in cylindrical coordinates. Due to the Galilean invariance of the Euler equations and of the GPR model, one can then simply superimpose a constant velocity field to this stationary vortex solution in order to get an unsteady version of the test problem. The vortex strength is chosen as $\varepsilon = 5$ and the perturbation of entropy $S = \frac{p}{\rho^\gamma}$ is assumed to be zero. For details of the setup, see Hu and Shu [175] and Dumbser et al. [59]. In this test we set the distortion field initially to $\boldsymbol{A} = \sqrt[3]{\rho}\,\boldsymbol{I}$, while the heat flux vector is initialized with $\boldsymbol{J} = 0$. As computational domain we choose $\Omega = [0; 10] \times [0; 10]$ with periodic boundary conditions. The reference solution for the GPR model in the stiff relaxation limit is given by the exact solution of the compressible Euler equatons, which is the time–shifted initial condition $\mathbf{Q}_e(\mathbf{x}, t) = \mathbf{Q}(\mathbf{x} - \boldsymbol{v}_c t, 0)$, where the convective mean velocity is $\boldsymbol{v}_c = (1, 1)$. We run this benchmark on a mesh sequence until the final time $t = 1.0$. The physical parameters of the GPR model are chosen as $\gamma = 1.4$, $c_v = 2.5$, $\rho_0 = 1$, $c_s = 0.5$, and $c_h = 1$. The volume fraction function is set to $\alpha = 1$ in the entire computational domain. The resulting numerical convergence rates obtained with ADER-DG schemes using polynomial approximation degrees from $N = M = 2$ to $N = M = 5$ are listed in **Table 1**, together with the chosen values for the effective viscosity μ and the effective heat conductivity coefficient κ. From **Table 1** one can observe that high order of convergence of the numerical method is achieved also in the stiff limit of the governing PDE system.

TABLE 1 | Experimental errors and order of accuracy at time $t = 1$ for the density ρ for ADER-DG schemes applied to the GPR model ($c_s = 0.5$, $\alpha = 1$) in the stiff relaxation limit ($\mu \ll 1, \kappa \ll 1$).

N_x	$\varepsilon(L_1)$	$\varepsilon(L_2)$	$\varepsilon(L_\infty)$	$\mathcal{O}(L_1)$	$\mathcal{O}(L_2)$	$\mathcal{O}(L_\infty)$
ADER-DG P_2P_2 ($\mu = \kappa = 10^{-6}$)						
20	9.4367E-03	2.2020E-03	2.1633E-03			
40	1.9524E-03	4.4971E-04	4.2688E-04	2.27	2.29	2.34
60	7.5180E-04	1.7366E-04	1.4796E-04	2.35	2.35	2.61
80	3.7171E-04	8.6643E-05	7.3988E-05	2.45	2.42	2.41
ADER-DG P_3P_3 ($\mu = \kappa = 10^{-6}$)						
10	1.7126E-02	4.0215E-03	3.6125E-03			
20	6.0405E-04	1.7468E-04	2.1212E-04	4.83	4.52	4.09
30	8.3413E-05	2.5019E-05	2.7576E-05	4.88	4.79	5.03
40	2.1079E-05	6.0168E-06	7.6291E-06	4.78	4.95	4.47
ADER DG P_4P_4 ($\mu = \kappa = 10^{-7}$)						
10	1.5539E-03	4.5965E-04	5.1665E-04			
20	4.3993E-05	1.0872E-05	1.0222E-05	5.14	5.40	5.66
25	1.8146E-05	4.4276E-06	4.1469E-06	3.97	4.03	4.04
30	8.6060E-06	2.1233E-06	1.9387E-06	4.09	4.03	4.17
ADER DG P_5P_5 ($\mu = \kappa = 10^{-7}$)						
5	1.1638E-02	1.1638E-02	1.8898E-03			
10	3.9653E-04	9.3717E-05	6.5319E-05	4.88	6.96	4.85
15	4.4638E-05	1.2572E-05	1.9056E-05	5.39	4.95	3.04
20	9.6136E-06	3.0120E-06	3.9881E-06	5.34	4.97	5.44

The reported errors are floating point numbers that have been obtained for numerical simulations carried out in double precision arithmetics.

4.2. Circular Explosion Problem in a Solid

In this Section, we simulate a circular explosion problem in an ideal elastic solid. We compare the results obtained with a third order ADER-WENO finite volume scheme on moving unstructured Voronoi meshes with possible topology changes, Gaburro et al. [77], with those obtained with a fourth order ADER discontinuous Galerkin finite element scheme on a very fine uniform Cartesian mesh composed of 512×512 elements, which will be taken as the reference solution for this benchmark. The computational domain is $\Omega = [-1, 1] \times [-1, 1]$ and the final simulation time is $t = 0.25$. We set $\alpha = 1$, $\boldsymbol{v} = \boldsymbol{0}$, $\boldsymbol{A} = \boldsymbol{I}$ and $\boldsymbol{J} = \boldsymbol{0}$ in the entire domain. For $r = \sqrt{x^2 + y^2} \leq 0.5$ the initial density and the initial pressure are set to $\rho = 1$ and $p = 1$, while in the rest of the domain we set $\rho = 0.1$ and $p = 10^{-3}$. The parameters of the GPR model are chosen as follows: $c_s = 0.2$, $c_h = 0$, $\tau_1 \rightarrow \infty$ (in order to model an elastic solid). We use the stiffened gas equation of state with $\gamma = 2$ and $p_0 = 0$. For the simulation on the moving Voronoi mesh, we employ a mesh with 82 919 control volumes. The computational results obtained with the unstructured ADER-WENO ALE scheme and those obtained with the high order Eulerian ADER-DG scheme are presented and compared with each other in **Figure 9**. We can note a very good agreement between the two results. The high quality of the ADER-WENO finite volume scheme on coarse grids is mainly due to the natural mesh refinement around the shock, which is typical for Lagrangian schemes. Furthermore, Lagrangian schemes are well-known to capture material interfaces and contact discontinuities very well, since the mesh is moving with the fluid and thus numerical dissipation at linear degenerate fields moving with the fluid velocity is significantly lower than with classical Eulerian schemes.

4.3. Rotor Test Problem

A second solid mechanics benchmark consists in the simulation of a plate on which a rotational impulse is initially impressed, in a circular region centered with respect to the computational domain. This *rotor* will initially move according to the rotational impulse, while emitting elastic waves which ultimately determine the formation of a set of concentric rings with alternating direction of rotation. The test is analogous to the rotor problem shown in Peshkov et al. [72], but with a weakened material in order to show stronger motion of the Voronoi grid.

The results of the third order ADER-WENO finite volume method on a moving Voronoi grid with variable connectivity, composed of 150 561 cells, are compared against a reference

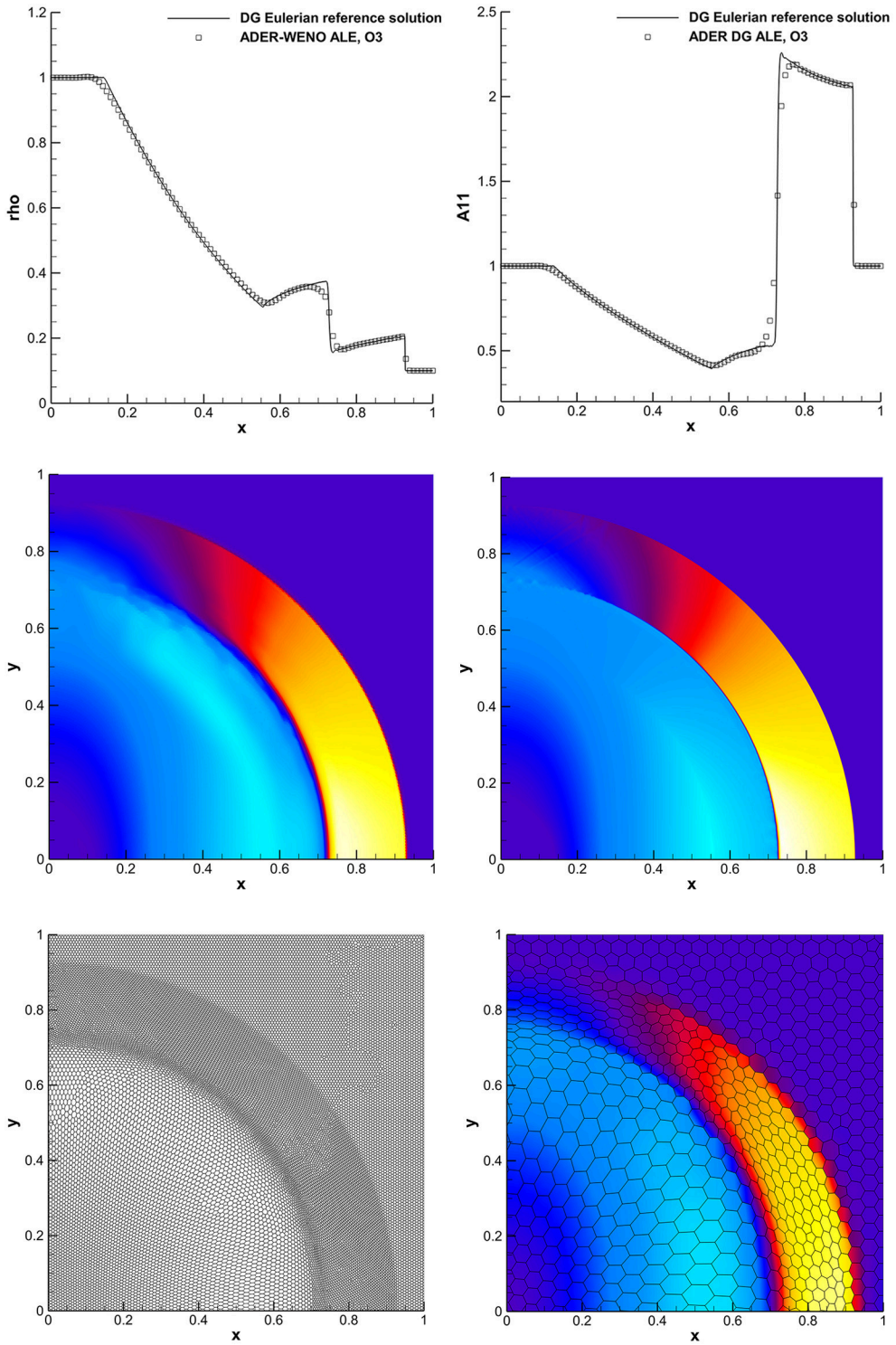


FIGURE 9 | Simulation results for the explosion problem obtained with a third order ADER-WENO ALE finite volume scheme on a moving Voronoi grid composed of 82 919 cells and with a fourth order ADER-DG scheme on a Cartesian grid of size $512^2 = 262\,144$ (4.2×10^6 DOF). In the top row, two cuts of the solution along the x-axis are shown; in the middle row, from the left, the solution for A_{11} obtained with the ADER-WENO ALE scheme and with the ADER-DG Eulerian scheme; in the bottom row, the Voronoi grid at the final simulation time and the results from the ADER-WENO ALE scheme on a coarser grid of 2 727 elements.

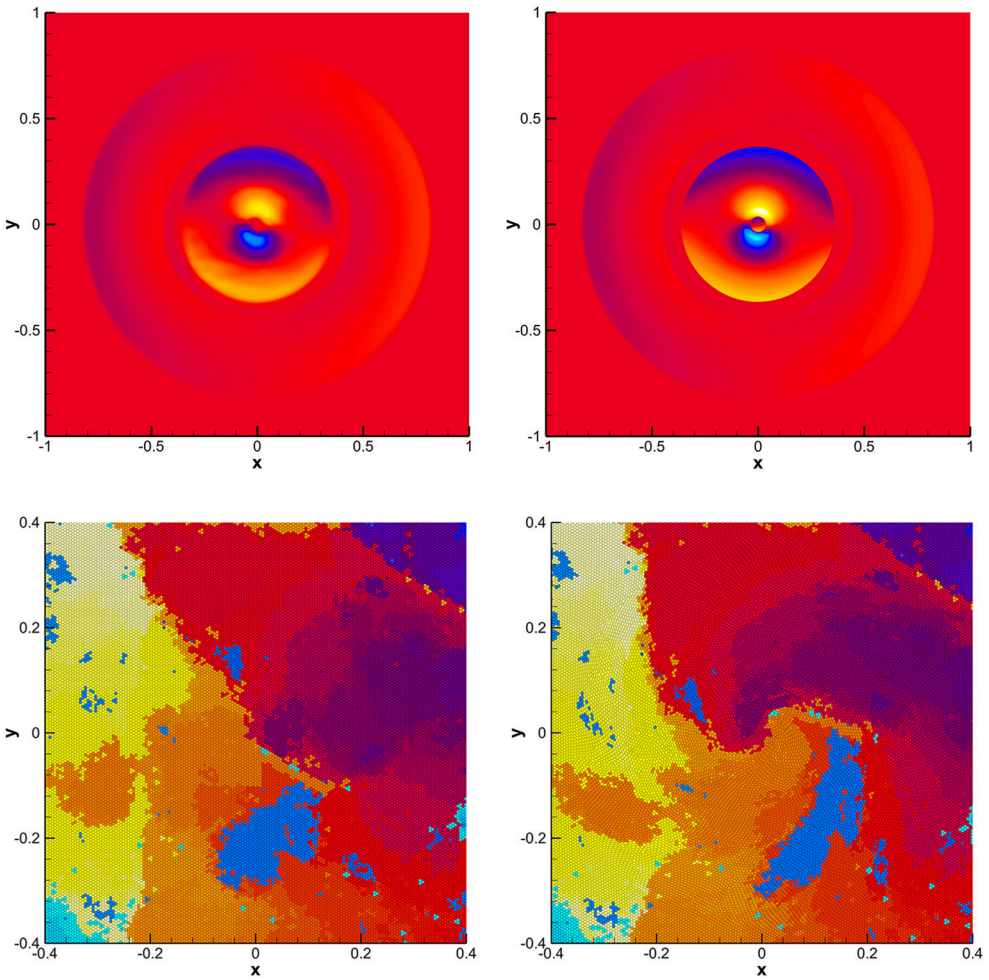


FIGURE 10 | Simulation results for the solid rotor problem obtained from a third order ADER-WENO ALE finite volume scheme on a moving Voronoi grid composed of 150 561 cells and with a fourth order ADER-DG scheme on a cartesian grid of size $512^2 = 262\,144$ (4.2×10^6 DOF). In the top row, the solutions for the u component of the velocity field are shown, on the left those obtained with the unstructured ADER-WENO ALE scheme on moving Voronoi meshes and on the right those of the ADER-DG scheme on a fixed Cartesian grid; in the bottom panels the cells are colored according to their mesh numbering to show the mesh motion between the beginning of the ALE simulation and the final time.

solution obtained with a fourth order ADER discontinuos Galerkin scheme on a very fine uniform Cartesian mesh counting 512×512 elements, for a total of over four million spatial degrees of freedom.

The computational domain is the square $\Omega = [-1, 1] \times [-1, 1]$ and the final simulation time is set to $t = 0.5$. With exception made for the velocity field, all variables are initially constant throughout the domain. Specifically we set $\alpha = 1$, $\rho = 1$, $p = 1$, $\boldsymbol{A} = \boldsymbol{I}$, $\boldsymbol{J} = \boldsymbol{0}$, while the velocity field is $\boldsymbol{v} = [-y/R, x/R, 0]$ if $r = \sqrt{x^2 + y^2} \leq R$, and $\boldsymbol{v} = \boldsymbol{0}$ otherwise, that is, outside of the circle of radius $R = 0.2$; this way, the initial tangential velocity at $r = R$ is one. The solid is taken to be elastic ($\tau_1 \rightarrow \infty$), heat wave propagation is neglected ($c_h = 0$), and the characteristic speed of shear waves is $c_s = 0.25$. The constitutive law is chosen to be the stiffened-gas EOS with $\gamma = 1.4$ and $p_0 = 0$. We can see in **Figure 10** that, although some of the finer features are lost (specifically the small central counterclockwise-rotating ring) due to the lower resolution of the finite volume method on a coarser grid, the shear waves travel outwards with the correct velocity and the moving Voronoi finite volume simulation can be said to be in agreement with the high resolution discontinuous Galerkin results. Also in **Figure 10**, it is shown that the central region of the computational grid has undergone significant motion but thanks to the absence of constraints on the connectivity between elements, the Voronoi control volumes have not been stretched excessively as would instead happen for a similar moving unstructured grid, but with fixed connectivity.

4.4. Elastic Vibrations of a Beryllium Plate

The first benchmark for our new diffuse interface version of the GPR model consists in the purely elastic vibrations of a beryllium plate, subject to an initial velocity distribution, see for example Sambasivan et al. [176], Maire et al. [177], Burton et al. [178],

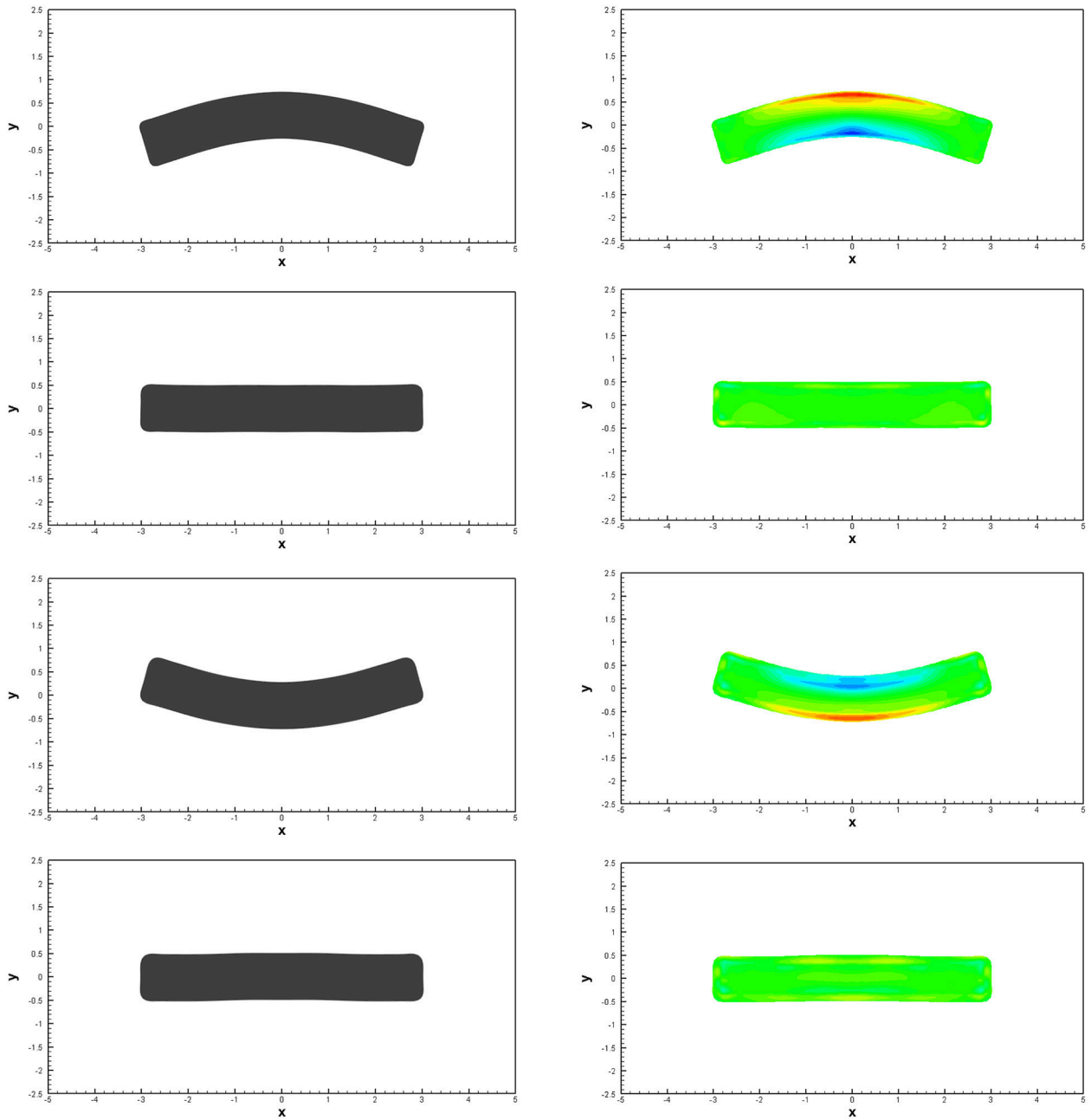


FIGURE 11 | Vibration of an elastic beryllium plate. Temporal evolution of the volume fraction function α (left) and of the pressure field (right) at times $t = 5$, $t = 14$, $t = 23$, and $t = 28$, from top to bottom.

Boscheri et al. [71], and Peshkov et al. [72] for a setup of the same test problem in the framework of Lagrangian and ALE schemes.

Unlike in the Lagrangian simulations, the computational domain considered here is *larger* and is set to $\Omega = [-5; 5] \times [-2.5; 2.5]$. The computational grid consists of 512×256 uniform Cartesian cells with a characteristic mesh size of about $h = 0.02$. We use a third order ADER-WENO finite volume scheme in the entire domain. The initial geometry of the beryllium bar is now simply defined by setting $\alpha(\mathbf{x}, 0) = 1 - \varepsilon$ inside the subdomain $\Omega_b = [-3, 3] \times [-0.5, 0.5]$, while the solid volume fraction function α is set to $\alpha(\mathbf{x}, 0) = \varepsilon$ elsewhere, with $\varepsilon = 5 \cdot 10^{-3}$. The initial velocity field inside Ω_b is imposed according to Burton et al. [178], Boscheri et al. [71], and Peshkov et al. [72] as

$$\boldsymbol{v}(\mathbf{x}) = \big(0, A\omega \left\{ C_1 \left(\sinh(\Omega(x+3)) + \sin(\Omega(x+3))\right) - S_1 \left(\cosh(\Omega(x+3)) + \cos(\Omega(x+3))\right)\right\}, 0\big), \quad (72)$$

with $\Omega = 0.7883401241$, $\omega = 0.2359739922$, $A = 0.004336850425$, $S_1 = 57.64552048$, and $C_1 = 56.53585154$, while we simply set $\boldsymbol{v} = \boldsymbol{0}$ outside Ω_b. For this test case we set $\varepsilon = 5 \cdot 10^{-3}$. The distortion field is initially set to $\boldsymbol{A} = \boldsymbol{I}$. The material properties of Beryllium in the Mie-Grüneisen equation of state are taken as follows: $\rho_0 = 1.845$, $c_0 = 1.287$, $c_s = 0.905$, $\Gamma = 1.11$, and $s_0 = 1.124$. We furthermore neglect heat conduction and set $c_h = 0$ and $\boldsymbol{J} = \boldsymbol{0}$.

Unlike in Lagrangian schemes, *no boundary conditions* need to be imposed on the surface of the bar. We simply use transmissive boundaries on $\partial\Omega$. The entire computational domain is initialized with the reference density for beryllium as $\rho(\mathbf{x}, 0) = \rho_0$, while the pressure is set to $p(\mathbf{x}, 0) = 0$. The distortion field is initialized with $\boldsymbol{A} = \boldsymbol{I}$. According to Burton et al. [178], the final time is set to $t_f = 53.25$ so that it corresponds approximately to two complete flexural periods. The simulations are carried out with a third order ADER-WENO scheme on two uniform Cartesian meshes composed of 256×128 and 512×256 elements, respectively.

For the fine grid simulation in **Figure 11**, we present the temporal evolution of the color contour map of the volume fraction function α, which represents the moving geometry of the bar. Here, dark gray color is used to indicate the regions with $\alpha > 0.5$ and white color is used for the regions of $\alpha < 0.5$. In the same figure, we also depict the pressure field in the region $\alpha > 0.5$ at times $t = 5$, $t = 14$, $t = 23$, and $t = 28$. These time instants cover approximately one flexural period. The time evolution of the vertical velocity component $v(0, 0, t)$ in the origin is depicted in **Figure 12**. For comparison, in the same figure we also show the results obtained on the coarse mesh for the same test problem with a fourth order ADER-DG scheme with second order TVD subcell finite volume limiter (red line).

Our computational results compare visually well against available reference solutions in the literature, see Sambasivan et al. [176], Maire et al. [177], Burton et al. [178], Boscheri et al. [71], and Peshkov et al. [72], which were all carried out with pure Lagrangian or Arbitrary-Lagrangian-Eulerian schemes on moving meshes, while here we use a diffuse interface approach on a fixed Cartesian grid.

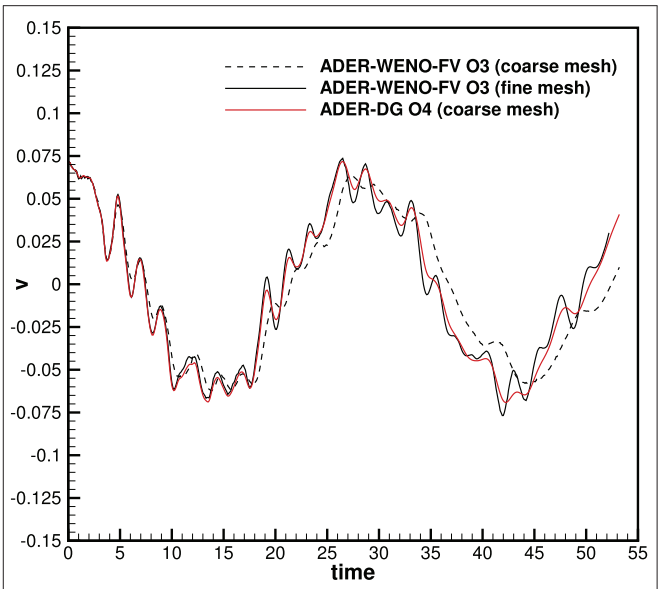


FIGURE 12 | Temporal evolution of the vertical velocity component $v(0, 0, t)$ obtained with a third order ADER-WENO scheme applied to the diffuse interface GPR model using two different mesh resolutions of 256×128 elements (coarse mesh) and 512×256 grid cells (fine mesh). For comparison, also a fourth order ADER-DG simulation on the coarse mesh is shown (red line).

4.5. Taylor Bar Impact Problem

So far, we have only considered *ideal elastic* material, i.e., the limit case $\tau_1 \rightarrow \infty$. In this section we consider also non-linear *elasto-plastic* material behavior. Following Barton et al. [179, 180], Boscheri et al. [71], and Peshkov et al. [72] we choose the relaxation time τ_1 as a non-linear function of an invariant of the stress tensor as follows:

$$\tau_1 = \tau_0 \left(\frac{\sigma_0}{\sigma} \right)^m, \tag{73}$$

where τ_0 is a constant characteristic relaxation time, σ_0 is the yield stress of the material and the von Mises stress σ is given by

$$\sigma = \sqrt{\frac{1}{2}((\sigma_{11} - \sigma_{22})^2 + (\sigma_{33} - \sigma_{11})^2 + (\sigma_{33} - \sigma_{22})^2 + 6(\sigma_{12}^2 + \sigma_{31}^2 + \sigma_{32}^2)} = \sqrt{\frac{3}{2}\, \mathring{\sigma}_{ij} \mathring{\sigma}_{ij}}. \tag{74}$$

In the formula (74) above, $\mathring{\sigma}_{ij} = \sigma_{ij} - \frac{1}{3}\sigma_{kk}\delta_{ij}$ is the stress deviator, i.e., the trace-free part of the stress tensor. The non-linear relaxation time (73) tends to zero for $\sigma \gg \sigma_0$, while it tends to infinity for $\sigma \ll \sigma_0$.

The Taylor bar impact problem is a classical benchmark for an elasto-plastic aluminium projectile that hits a rigid solid wall, see Sambasivan et al. [176], Maire et al. [177], Dobrev et al. [181], and Boscheri et al. [71]. In this work the computational domain under consideration is $\Omega = [0, 600] \times [-150, +150]$. The aluminium bar is initially located in the region $\Omega_b = [0, 500] \times [-50, +50]$, where we set $\alpha = 1 - \varepsilon$, while in the rest of the computational domain we set $\alpha = \varepsilon$, with $\varepsilon = 1 \cdot 10^{-2}$.

The aluminium bar is described by the Mie-Grüneisen equation of state with parameters $\rho_0 = 2.785$, $c_0 = 0.533$, $c_s = 0.305$, $\Gamma = 2$, and $s = 1.338$. The yield stress of aluminium is set to $\sigma_0 = 0.003$.

The projectile is initially moving with velocity $\boldsymbol{v} = (-0.015, 0)$ toward a wall located at $x = 0$. This velocity field is imposed within the subregion Ω_b, while in the rest of the domain we set $\boldsymbol{v} = \boldsymbol{0}$. The remaining initial conditions are chosen as $\rho = \rho_0$, $p = p_0$, $\boldsymbol{A} = \boldsymbol{I}$, $\boldsymbol{J} = \boldsymbol{0}$ and with the parameters $\tau_0 = 1$ and $m = 20$ for the computation of the relaxation time (73). Unlike in Lagrangian schemes, we do not need to set any boundary conditions on the free surface of the moving bar. We only apply reflective slip wall boundary conditions on the wall in $x = 0$. According to Maire et al. [177], Dobrev et al. [181], and Boscheri et al. [71] the final time of the simulation is $t = 5,000$. The computational domain is discretized on a regular Cartesian grid composed of 512×256 elements using a third order ADER-WENO finite volume scheme. As in Boscheri et al. [71] we

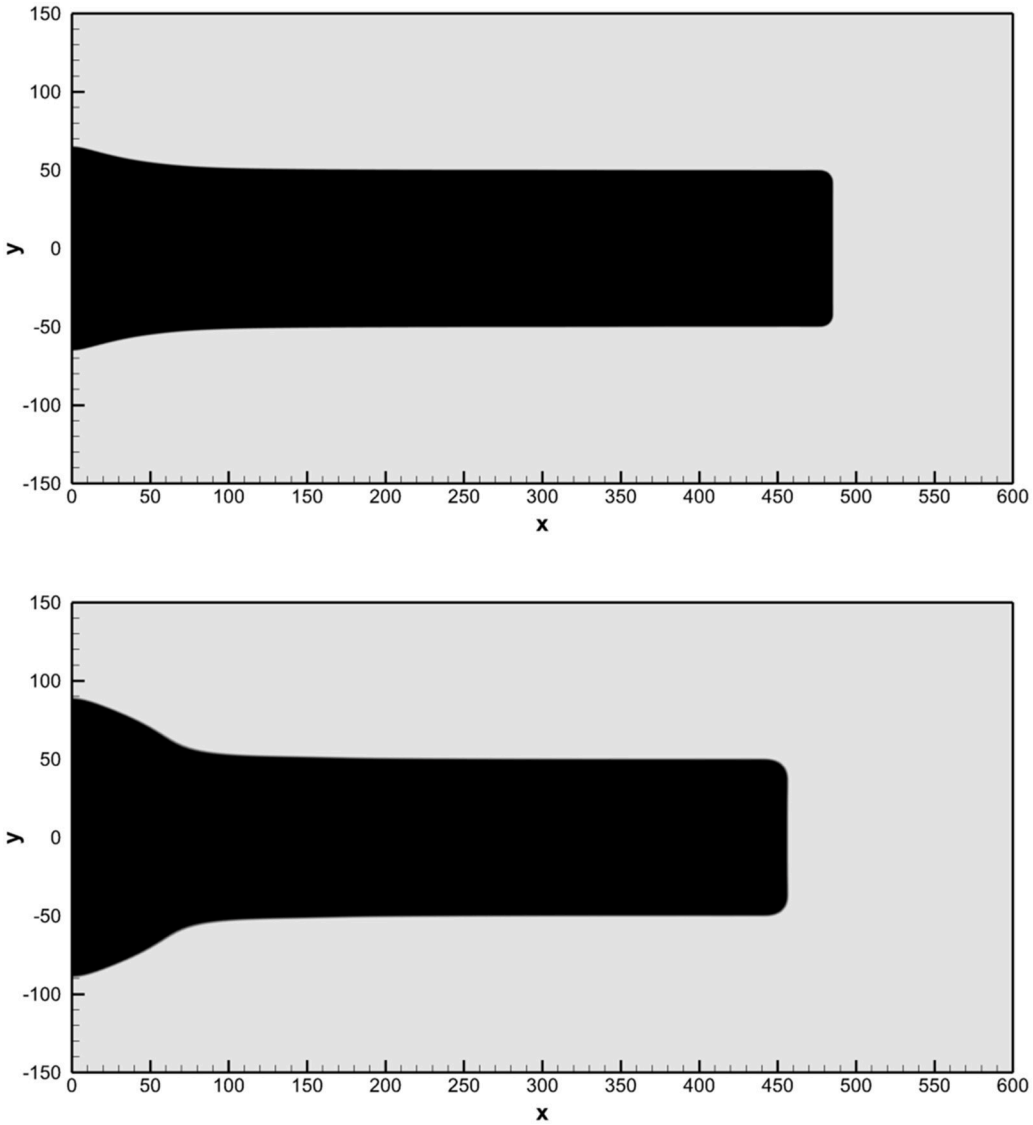


FIGURE 13 | Geometry of the Taylor bar at time $t = 1,000$ **(top)** and at the final time $t = 5,000$ **(bottom)** obtained with a third order ADER-WENO finite volume scheme applied to the diffuse interface GPR model. We plot the contour colors of the volume fraction function α, where black regions denote $\alpha > 0.5$ and white regions $\alpha < 0.5$.

employ a classical source splitting for the treatment of the stiff sources that arise in the regions of plastic deformations, i.e., when $\sigma \gg \sigma_0$. In **Figure 13**, we show the computational results at $t = 1000$ and at the final time $t = 5,000$. The obtained solution is in agreement with the results presented in Maire et al. [177], Boscheri et al. [71], and Peshkov et al. [72]. At time $t = 5,000$, we measure a final length of the projectile of $L_f = 456$, which fits the results achieved in Maire et al. [177] and Boscheri et al. [71] up to 2%.

4.6. Dambreak Problem

In this last section on numerical test problems, we solve a two-dimensional dambreak problem with different relaxation times in order to show the entire range of potential applications of the GPR model. For this purpose, we also activate the gravity source term, setting the gravity vector to $g = (0, -g)$ with $g = 9.81$. The computational domain is chosen as $\Omega = [0, 4] \times [0, 2]$ and is discretized with a fourth order ADER discontinuous Galerkin finite element scheme with polynomial approximation degree $N = 3$ and *a posteriori* subcell TVD finite volume limiter. Computations are run on a uniform Cartesian mesh composed of 128×64 elements until the final time $t = 0.5$. The initial condition is chosen as follows: we set $\rho = \rho_0$, $\boldsymbol{v} = \boldsymbol{0}$, $\boldsymbol{A} = \boldsymbol{I}$ and $\boldsymbol{J} = \boldsymbol{0}$ in the entire computational domain. We impose the slip boundary condition on the bottom. In the subdomain $\Omega_d = [0, 2] \times [0, 1]$, we set $\alpha = 1 - \varepsilon$, and $p = \rho_0 g(y - 1)$, while in the rest of the domain we set $\alpha = \varepsilon$ and $p = 0$. In this test problem we set $\varepsilon = 10^{-2}$ and use a stiffened gas

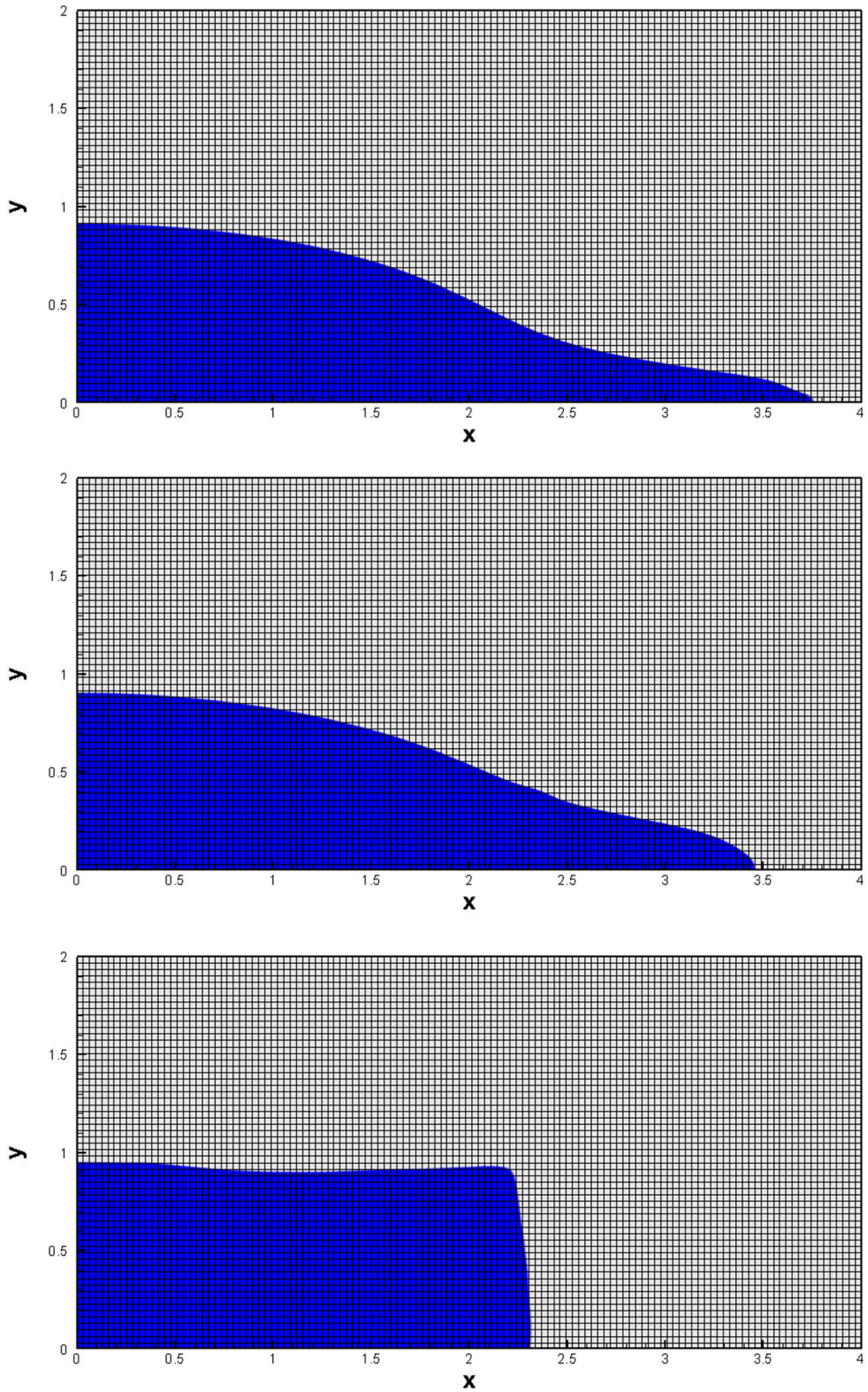


FIGURE 14 | Dambreak problem at $t = 0.5$, simulated with a fourth order ADER-DG scheme using different relaxation times. **(Top)** Low viscosity fluid (stiff relaxation limit) with $\nu = 10^{-3}$. **(Center)** High viscosity fluid with $\nu = 10^{-1}$. **(Bottom)** Ideal elastic solid ($\tau_1 \rightarrow \infty$) with low shear resistance.

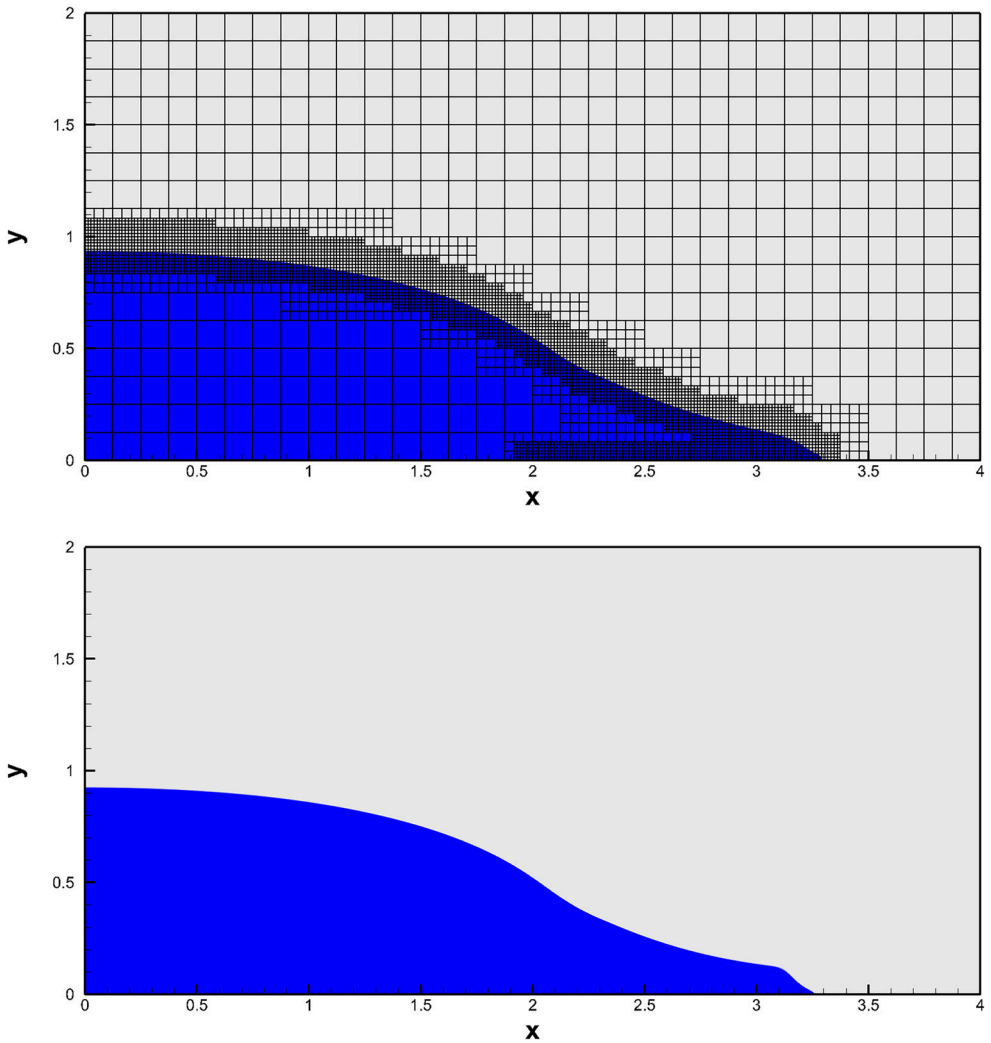


FIGURE 15 | Dambreak problem at $t = 0.4$, simulated with a fourth order ADER-DG scheme using a space-time adaptive Cartesian AMR mesh applied to the GPR model with with $\nu = 10^{-3}$ **(Top)**, and reference solution, computed with a third order ADER-WENO finite volume scheme on a very fine uniform Cartesian grid, solving the inviscid and barotropic reduced Baer-Nunziato approach presented in [106, 107] **(Bottom)**.

equation of state with parameters $\rho_0 = 1,000$, $p_0 = 5 \times 10^4$, $\gamma = 2$, $c_h = 0$ and a shear sound speed $c_s = 6$. Simulations are run in three different regimes, only characterized by a different choice of the strain relaxation time τ_1. In the first simulation, we set τ_1 so that a kinematic viscosity $\nu = \mu/\rho_0 = 10^{-3}$ is reached in the stiff relaxation limit, i.e., the GPR model in this case describes an almost inviscid fluid. In the second simulation we choose τ_1 so that $\nu = 0.1$, i.e., a high viscosity Newtonian fluid behavior is reached. In the last simulation we set $\tau_1 \rightarrow \infty$, i.e., the strain relaxation term is switched off so that an ideal elastic solid with low shear resistance is described, similar to a jelly-type medium. In all cases, we apply solid slip wall boundary conditions on the left and on the right of the computational domain, while on the right and upper boundary, transmissive boundary conditions are set. The temporal evolution of the volume fraction function α, together with the coarse mesh used in this simulation, are depicted in **Figure 14**. The results for the almost inviscid fluid agree qualitatively well with those shown in Ferrari et al. [182], Dumbser et al. [106], and Gaburro et al. [107] for non-hydrostatic dambreak problems. In order to corroborate this statement quantitatively, we now repeat the simulation with $\nu = 10^{-3}$ using a fourth order ADER-DG scheme on a coarse AMR grid composed of only 32×16 elements on the level zero grid. We then apply two levels of AMR refinement with refinement factor $\mathfrak{r} = 3$, i.e., we employ a general space-tree, rather than a simple quad-tree. We note that the simulations on the AMR grid are run in combination with time-accurate local time stepping (LTS), which is trivial to implement in high order ADER-DG and ADER-FV schemes, due to their fully-discrete one-step nature. For details on LTS, see Dumbser et al. [37, 54],Dumbser [64] and Gaburro et al. [65]. As a reference solution of this almost inviscid flow problem, we solve the reduced barotropic and

inviscid Baer-Nunziato model introduced in Dumbser [106] and Gaburro et al. [107], using a third order ADER-WENO finite volume scheme on a very fine uniform Cartesian grid composed of 1024×512 elements. The direct comparison of the two simulations at time $t = 0.4$ is shown in **Figure 15**. Overall we can indeed note an excellent agreement between the behavior of the diffuse interface GPR model in the stiff relaxation limit and the weakly compressible inviscid non-hydrostatic free surface flow model of Dumbser [106] and Gaburro et al. [107].

5. CONCLUSIONS AND OUTLOOK

In the first part of this paper we have provided a review of the ADER approach, whose development started about 20 years ago with the seminal works of Toro et al. [20] Millington et al. [19], Titarev and Toro [29], and Toro and Titarev [28] in the context of approximate solvers for the generalized Riemann problem (GPR). The ADER method provides *fully discrete* explicit one-step schemes that are in principle arbitrary high order accurate in both space and time. The most recent developments include ADER schemes for stiff source terms, as well as ADER finite volume and discontinuous Galerkin finite element schemes on fixed and moving meshes, which are all based on a space-time predictor-corrector approach. The fact that ADER schemes are fully discrete makes the implementation of time accurate local time stepping (LTS) particularly simple, both on adaptive Cartesian AMR meshes [54], as well as in the context of Lagrangian schemes on moving grids [64, 65]. The fully discrete space-time formulation also allows the treatment of topology changes during one time step in a very natural way [77]. In the second part of the paper we have then shown several applications of high order ADER finite volume and discontinuous Galerkin finite element schemes to the novel unified hyperbolic model of continuum mechanics (GPR model) proposed by Godunov, Peshkov and Romenski [56, 57, 59]. The presented test problems cover the entire range of continuum mechanics, from ideal elastic solids over plastic solids to viscous fluids. The use of a diffuse interface approach allows also to simulate moving boundary problems on fixed Cartesian meshes. Future developments will concern the extension of the mathematical model to non-Newtonian fluids [183] and to free surface flows with surface tension, see Schmidmayer et al. [184] and Chiocchetti et al. [185], as well as to the conservative multi-phase model of Romenski et al. [186, 187]. In future work we will also consider the use of novel all speed schemes [188] and semi-implicit space-time discontinuous Galerkin finite element schemes [189–191] for the diffuse interface version of the GPR model used in this paper.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

The governing PDE system was developed by IP. The numerical method and the computer codes were developed by MD, EG, and SC. The test problems were computed by MD, EG, and SC. The analysis of the method was performed by SB. All authors discussed the results and contributed to the final manuscript.

FUNDING

The research presented in this paper has been financed by the European Union's Horizon 2020 Research and Innovation Programme under the project *ExaHyPE*, grant agreement number no. 671698 (call FET-HPC-1-2014). SB has also received funding by INdAM (*Istituto Nazionale di Alta Matematica*, Italy) under a Post-doctoral grant of the research project *Progetto premiale FOE 2014-SIES*. SC acknowledges the financial support received by the Deutsche Forschungsgemeinschaft (DFG) under the project *Droplet Interaction Technologies (DROPIT)*, grant no. GRK 2160/1. MD also acknowledges the financial support received from the Italian Ministry of Education, University and Research (MIUR) in the frame of the Departments of Excellence Initiative 2018–2022 attributed to DICAM of the University of Trento (grant L. 232/2016) and in the frame of the PRIN 2017 project. MD has also received funding from the University of Trento via the *Strategic Initiative Modeling and Simulation*. EG has also been financed by a national mobility grant for young researchers in Italy, funded by GNCS-INdAM and acknowledges the support given by the University of Trento through the *UniTN Starting Grant* initiative.

REFERENCES

1. Godunov SK. Finite difference methods for the computation of discontinuous solutions of the equations of fluid dynamics. *Math USSR*. (1959) **47**:271–306.
2. Lax P, Wendroff B. Systems of conservation laws. *Commun Pure Appl Math*. (1960) **13**:217–37. doi: 10.1002/cpa.3160130205
3. Kolgan VP. Application of the minimum-derivative principle in the construction of finite-difference schemes for numerical analysis of discontinuous solutions in gas dynamics. *Trans Central Aerohydrodyn Inst*. (1972) **3**:68–77.
4. Harten A. High resolution schemes for hyperbolic conservation laws. *J Comput Phys*. (1983) **49**:357–93. doi: 10.1016/0021-9991(83)90136-5
5. Sweby PK. High resolution TVD schemes using flux limiters. *Lect Appl Math*. (1985) **22**:289–309.
6. Gottlieb S, Shu C. Total variation diminishing Runge-Kutta schemes. *Math Comput*. (1998) **67**:73–85. doi: 10.1090/S0025-5718-98-00913-2
7. van Leer B. Towards the ultimate conservative difference scheme V: a second order sequel to Godunov's Method. *J Comput Phys*. (1979) **32**:101–36. doi: 10.1016/0021-9991(79)90145-1
8. van Leer B. On the relationship between the upwind-differencing schemes of Godunov, Engquist-Osher and Roe. *SIAM J Sci Stat Comput*. (1985) **5**:1–20. doi: 10.1137/0905001
9. Toro E. *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction*. Berlin; Heidelberg: Springer-Verlag (2009).
10. Harten A, Osher S. Uniformly high-order accurate nonoscillatory schemes I. *SIAM J Num Anal*. (1987) **24**:279–309. doi: 10.1137/0724022
11. Harten A, Engquist B, Osher S, Chakravarthy S. Uniformly high order accurate essentially non-oscillatory schemes III. *J Comput Phys*. (1987) **71**:231–303. doi: 10.1016/0021-9991(87)90031-3

12. Shu C. *Essentially Non-Oscillatory and Weighted Essentially Non-Oscillatory Schemes for Hyperbolic Conservation Laws*. NASA/CR-97-206253 ICASE Report No97-65. (1997).
13. Castro CC, Toro EF. Solvers for the high-order Riemann problem for hyperbolic balance laws. *J Comput Phys*. (2008) **227**:2481–513. doi: 10.1016/j.jcp.2007.11.013
14. Berthon C. Why the MUSCL-Hancock scheme is L1-stable. *Numer Math*. (2006) **104**:27–46. doi: 10.1007/s00211-006-0007-4
15. Ben-Artzi M, Falcovitz J. A second-order Godunov-type scheme for compressible fluid dynamics. *J Comput Phys*. (1984) **55**:1–32. doi: 10.1016/0021-9991(84)90013-5
16. LeFloch P, Tatsien L. A global asymptotic expansion for the solution of the generalized Riemann problem. *Ann Inst Henri Poincaré (C) Analyse Non Linéaire*. (1991) **3**:321–40. doi: 10.3233/ASY-1991-3404
17. Ben-Artzi M, Li J, Warnecke G. A direct Eulerian GRP scheme for Compressible Fluid Flows. *J Comput Phys*. (2006) **218**:19–43. doi: 10.1016/j.jcp.2006.01.044
18. Han E, Li J, Tang H. An adaptive GRP scheme for compressible fluid flows. *J Comput Phys*. (2010) **229**:1448–66. doi: 10.1016/j.jcp.2009.10.038
19. Millington R, Toro E, Nejad L. *Arbitrary High Order Methods for Conservation Laws I: The One Dimensional Scalar Case*. Manchester Metropolitan University; Department of Computing and Mathematics (1999).
20. Toro E, Millington R, Nejad L. Towards very high order Godunov schemes. In: Toro E, editor. *Godunov Methods. Theory and Applications*. Boston, MA: Springer (2001). p. 905–38.
21. Toro EF. A weighted average flux method for hyperbolic conservation laws. *Proc R Soc Lond A Math Phys Eng Sci*. (1989) **423**:401–18. doi: 10.1098/rspa.1989.0062
22. Billett S, Toro E. On the accuracy and stability of explicit Schemes for Multidimensional linear homogeneous Advection Equations. *J Comput Phys*. (1997) **131**:247–50. doi: 10.1006/jcph.1996.5610
23. van Leer B. Towards the Ultimate Conservative Difference Scheme II: monotonicity and conservation combined in a second order scheme. *J Comput Phys*. (1974) **14**:361–70. doi: 10.1016/0021-9991(74)90019-9
24. Colella P. A direct Eulerian MUSCL scheme for gas dynamics. *SIAM J Sci Stat Comput*. (1985) **6**:104–17. doi: 10.1137/0906009
25. Ben-Artzi M, Falcovitz J. *Generalized Riemann Problems in Computational Fluid Dynamics*. No. 11 in Cambridge Monographs on Applied and Computational Mathematics. Cambridge: Cambridge University Press (2003).
26. Schwartzkopff T, Munz C, Toro E. ADER: a high order approach for linear hyperbolic systems in 2D. *J Sci Comput*. (2002) **17**:231–40. doi: 10.1023/A:1015160900410
27. Schwartzkopff T, Dumbser M, Munz C. Fast high order ADER schemes for linear hyperbolic equations. *J Comput Phys*. (2004) **197**:532–9. doi: 10.1016/j.jcp.2003.12.007
28. Toro E, Titarev V. Solution of the generalized Riemann problem for advection-reaction equations. *Proc R Soc Lond*. (2002) **458**:271–81. doi: 10.1098/rspa.2001.0926
29. Titarev V, Toro E. ADER: arbitrary high order Godunov approach. *J Sci Comput*. (2002) **17**:609–18. doi: 10.1023/A:1015126814947
30. Käser MA. *Adaptive Methods for the Numerical Simulation of Transport Processes*. Technische Universität München (2003).
31. Käser M, Iske A. Adaptive ADER schemes for the solution of scalar non-linear hyperbolic problems. *J Comput Phys*. (2005) **205**:489–508. doi: 10.1016/j.jcp.2004.11.015
32. Dumbser M, Käser M, Titarev VA, Toro EF. Quadrature-free non-oscillatory finite volume schemes on unstructured meshes for nonlinear hyperbolic systems. *J Comput Phys*. (2007) **226**:204–43. doi: 10.1016/j.jcp.2007.04.004
33. Qiu J, Dumbser M, Shu C. The discontinuous Galerkin Method with Lax-Wendroff type time discretizations. *Comput Methods Appl Mech Eng*. (2005) **194**:4528–43. doi: 10.1016/j.cma.2004.11.007
34. Dumbser M, Munz C. Building blocks for arbitrary high order discontinuous Galerkin schemes. *J Sci Comput*. (2006) **27**:215–30. doi: 10.1007/s10915-005-9025-0
35. Gassner G, Dumbser M, Hindenlang F, Munz CD. Explicit one–step time discretizations for discontinuous Galerkin and finite volume schemes based on local predictors. *J Comput Phys*. (2011) **230**:4232–47. doi: 10.1016/j.jcp.2010.10.024
36. Fambri F, Dumbser M, Köppel S, Rezzolla L, Zanotti O. ADER discontinuous Galerkin schemes for general-relativistic ideal magnetohydrodynamics. *Mon Notices R Astron Soc*. (2018) **477**:4543–64. doi: 10.1093/mnras/sty734
37. Dumbser M, Käser M, Toro EF. An arbitrary high order discontinuous Galerkin method for elastic waves on unstructured meshes V: local time stepping and *p*-adaptivity. *Geophys J Int*. (2007) **171**:695–717. doi: 10.1111/j.1365-246X.2007.03427.x
38. Dumbser M, Enaux C, Toro EF. Finite volume schemes of very high order of accuracy for stiff hyperbolic balance laws. *J Comput Phys*. (2008) **227**:3971–4001. doi: 10.1016/j.jcp.2007.12.005
39. van der Vegt JJW, van der Ven H. Space–time discontinuous Galerkin finite element method with dynamic grid motion for inviscid compressible flows I. General formulation. *J Comput Phys*. (2002) **182**:546–85. doi: 10.1006/jcph.2002.7185
40. van der Ven H, van der Vegt JJW. Space–time discontinuous Galerkin finite element method with dynamic grid motion for inviscid compressible flows II. Efficient flux quadrature. *Comput Methods Appl Mech Eng*. (2002) **191**:4747–80. doi: 10.1016/S0045-7825(02)00403-6
41. Hidalgo A, Dumbser M. ADER schemes for nonlinear systems of stiff advection-diffusion-reaction equations. *J Sci Comput*. (2011) **48**:173–89. doi: 10.1007/s10915-010-9426-6
42. Dumbser M, Balsara D, Toro EF, Munz CD. A unified framework for the construction of one-step finite–volume and discontinuous Galerkin schemes. *J Comput Phys*. (2008) **227**:8209–53. doi: 10.1016/j.jcp.2008.05.025
43. Luo H, Luo L, Nourgaliev R, Mousseau VA, Dinh N. A reconstructed discontinuous Galerkin method for the compressible Navier-Stokes equations on arbitrary grids. *J Comput Phys*. (2010) **229**:6961–78. doi: 10.1016/j.jcp.2010.05.033
44. Luo H, Xia Y, Spiegel S, Nourgaliev R, Jiang Z. A reconstructed discontinuous Galerkin method based on a Hierarchical WENO reconstruction for compressible flows on tetrahedral grids . *J Comput Phys*. (2013) **236**:477–92. doi: 10.1016/j.jcp.2012.11.026
45. Dumbser M, Zanotti O. Very high order PNPM schemes on unstructured meshes for the resistive relativistic MHD equations. *J Comput Phys*. (2009) **228**:6991–7006. doi: 10.1016/j.jcp.2009.06.009
46. Dumbser M, Castro M, Parés C, Toro EF. ADER schemes on unstructured meshes for non-conservative hyperbolic systems: applications to geophysical flows. *Comput Fluids*. (2009) **38**:1731–48. doi: 10.1016/j.compfluid.2009.03.008
47. Dumbser M, Hidalgo A, Castro M, Parés C, Toro EF. FORCE schemes on unstructured meshes II: non-conservative hyperbolic systems. *Comput Methods Appl Mech Eng*. (2010) **199**:625–47. doi: 10.1016/j.cma.2009.10.016
48. Dumbser M. Arbitrary high order PNPM schemes on unstructured meshes for the compressible Navier–Stokes equations. *Comput Fluids*. (2010) **39**:60–76. doi: 10.1016/j.compfluid.2009.07.003
49. Balsara DS, Dumbser M, Abgrall R. Multidimensional HLLC riemann solver for unstructured meshes - With application to Euler and MHD flows. *J Comput Phys*. (2014) **261**:172–208. doi: 10.1016/j.jcp.2013.12.029
50. Balsara DS, Dumbser M. Divergence-free MHD on unstructured meshes using high order finite volume schemes based on multidimensional Riemann solvers. *J Comput Phys*. (2015) **299**:687–715. doi: 10.1016/j.jcp.2015.07.012
51. Zanotti O, Fambri F, Dumbser M. Solving the relativistic magnetohydrodynamics equations with ADER discontinuous Galerkin methods, a posteriori subcell limiting and adaptive mesh refinement. *Mon Not R Astron Soc*. (2015) **452**:3010–29. doi: 10.1093/mnras/stv1510
52. Dumbser M, Guercilena F, Köppel S, Rezzolla L, Zanotti O. Conformal and covariant Z4 formulation of the Einstein equations: strongly hyperbolic first–order reduction and solution with discontinuous Galerkin schemes. *Phys Rev D*. (2018) **97**:084053. doi: 10.1103/PhysRevD.97.084053
53. Dumbser M, Fambri F, Gaburro E, Reinarz A. On GLM curl cleaning for a first order reduction of the CCZ4 formulation of the Einstein field equations. *J Comput Phys*. (2020) **404**:109088. doi: 10.1016/j.jcp.2019.109088
54. Dumbser M, Zanotti O, Hidalgo A, Balsara DS. ADER-WENO finite volume schemes with space-time adaptive mesh refinement. *J Comput Phys*. (2013) **248**:257–86. doi: 10.1016/j.jcp.2013.04.017

55. Dumbser M, Hidalgo A, Zanotti O. High order space-time adaptive ADER-WENO finite volume schemes for non-conservative hyperbolic systems. *Comput Methods Appl Mech Eng*. (2014) **268**:359–87. doi: 10.1016/j.cma.2013.09.022
56. Godunov SK, Romenskii EI. Nonstationary equations of nonlinear elasticity theory in eulerian coordinates. *J Appl Mech Tech Phys*. (1972) **13**:868–84. doi: 10.1007/BF01200547
57. Peshkov I, Romenski E. A hyperbolic model for viscous Newtonian flows. *Continuum Mech Thermodyn*. (2016) **28**:85–104. doi: 10.1007/s00161-014-0401-6
58. Godunov SK, Romenski EI. *Elements of Continuum Mechanics and Conservation Laws*. Boston, MA: Kluwer Academic/Plenum Publishers (2003).
59. Dumbser M, Peshkov I, Romenski E, Zanotti O. High order ADER schemes for a unified first order hyperbolic formulation of continuum mechanics: viscous heat-conducting fluids and elastic solids. *J Comput Phys*. (2016) **314**:824–62. doi: 10.1016/j.jcp.2016.02.015
60. Dumbser M, Peshkov I, Romenski E, Zanotti O. High order ADER schemes for a unified first order hyperbolic formulation of Newtonian continuum mechanics coupled with electro-dynamics. *J Comput Phys*. (2017) **348**:298–342. doi: 10.1016/j.jcp.2017.07.020
61. Dumbser M, Fambri F, Tavelli M, Bader M, Weinzierl T. Efficient implementation of ADER discontinuous Galerkin schemes for a scalable hyperbolic PDE engine. *Axioms*. (2018) **7**:63. doi: 10.3390/axioms7030063
62. Boscheri W, Dumbser M. Arbitrary–Lagrangian–Eulerian One–step WENO finite volume schemes on unstructured triangular meshes. *Commun Comput Phys*. (2013) **14**:1174–206. doi: 10.4208/cicp.181012.010313a
63. Boscheri W, Dumbser M. A direct Arbitrary-Lagrangian-Eulerian ADER-WENO finite volume scheme on unstructured tetrahedral meshes for conservative and non-conservative hyperbolic systems in 3D. *J Comput Phys*. (2014) **275**:484–523. doi: 10.1016/j.jcp.2014.06.059
64. Dumbser M. Arbitrary-Lagrangian-Eulerian ADER-WENO finite volume schemes with time-accurate local time stepping for hyperbolic conservation laws. *Comput Methods Appl Mech Eng*. (2014) **280**:57–83. doi: 10.1016/j.cma.2014.07.019
65. Boscheri W, Dumbser M, Zanotti O. High order cell-centered Lagrangian-type finite volume schemes with time-accurate local time stepping on unstructured triangular meshes. *J Comput Phys*. (2014) **291**:120–50. doi: 10.1016/j.jcp.2015.02.052
66. Boscheri W, Balsara DS, Dumbser M. Lagrangian ADER-WENO finite volume schemes on unstructured triangular meshes based On Genuinely Multidimensional HLL Riemann Solvers. *J Comput Phys*. (2014) **267**:112–38. doi: 10.1016/j.jcp.2014.02.023
67. Bonazzoli M, Gaburro E, Dolean V, Rapetti F. High order edge finite element approximations for the time-harmonic Maxwell's equations. In: *2014 IEEE Conference on Antenna Measurements and Applications (CAMA)*. Antibes Juan-les-Pins: IEEE (2014). p. 1–4.
68. Boscheri W, Dumbser M, Balsara DS. High order Lagrangian ADER-WENO schemes on unstructured meshes – Application of several node solvers to hydrodynamics and Magnetohydrodynamics. *Int J Numer Methods Fluids*. (2014) **76**:737–78. doi: 10.1002/fld.3947
69. Boscheri W, Dumbser M. An efficient quadrature-free formulation for high order Arbitrary-Lagrangian-Eulerian ADER-WENO finite volume schemes on unstructured meshes. *J Sci Comput*. (2016) **66**:240–74. doi: 10.1007/s10915-015-0019-2
70. Boscheri W, Dumbser M. High order accurate direct Arbitrary-Lagrangian-Eulerian ADER-WENO finite volume schemes on moving curvilinear unstructured meshes. *Comput Fluids*. (2016) **136**:48–66. doi: 10.1016/j.compfluid.2016.05.020
71. Boscheri W, Dumbser M, Loubere R. Cell centered direct Arbitrary-Lagrangian-Eulerian ADER-WENO finite volume schemes for nonlinear hyperelasticity. *Comput Fluids*. (2016) **134–35**:111–29. doi: 10.1016/j.compfluid.2016.05.004
72. Peshkov I, Boscheri W, Loubère R, Romenski E, Dumbser M. Theoretical and numerical comparison of hyperelastic and hypoelastic formulations for Eulerian non–linear elastoplasticity. *J Comput Phys*. (2019) **387**:481–521. doi: 10.1016/j.jcp.2019.02.039
73. Gaburro E, Dumbser M, Castro M. Direct Arbitrary-Lagrangian-Eulerian finite volume schemes on moving nonconforming unstructured meshes. *Comput Fluids*. (2017) **159**:254–75. doi: 10.1016/j.compfluid.2017.09.022
74. Gaburro E, Castro M, Dumbser M. Well balanced Arbitrary-Lagrangian-Eulerian finite volume schemes on moving nonconforming meshes for the Euler equations of gasdynamics with gravity. *Mon Notices R Astron Soc*. (2018) **477**:2251–75. doi: 10.1093/mnras/sty542
75. Springel V. E pur si muove: galilean-invariant cosmological hydrodynamical simulations on a moving mesh. *Mon Notices R Astron Soc*. (2010) **401**:791–851. doi: 10.1111/j.1365-2966.2009.15715.x
76. Gaburro E. A unified framework for the solution of hyperbolic pde systems using high order direct arbitrary-lagrangian-eulerian schemes on moving unstructured meshes with topology change. *Arch Comput Methods Eng*. (2020). doi: 10.1007/s11831-020-09411-7
77. Gaburro E, Boscheri W, Chiocchetti S, Klingenberg C, Springel V, Dumbser M. High order direct Arbitrary-Lagrangian-Eulerian schemes on moving Voronoi meshes with topology changes. *J Comput Phys*. (2020) 109167. doi: 10.1016/j.jcp.2019.109167
78. Dumbser M, Zanotti O, Loubère R, Diot S. A posteriori subcell limiting of the discontinuous Galerkin finite element method for hyperbolic conservation laws. *J Comput Phys*. (2014) **278**:47–75. doi: 10.1016/j.jcp.2014.08.009
79. Loubère R, Dumbser M, Diot S. A new family of high order unstructured MOOD and ADER finite volume schemes for multidimensional systems of hyperbolic conservation laws. *Commun Comput Phys*. (2014) **16**:718–63. doi: 10.4208/cicp.181113.140314a
80. Zanotti O, Fambri F, Dumbser M, Hidalgo A. Space–time adaptive ADER discontinuous Galerkin finite element schemes with a posteriori sub–cell finite volume limiting. *Comput Fluids*. (2015) **118**:204–24. doi: 10.1016/j.compfluid.2015.06.020
81. Dumbser M, Loubère R. A simple robust and accurate a posteriori sub–cell finite volume limiter for the discontinuous Galerkin method on unstructured meshes. *J Comput Phys*. (2016) **319**:163–99. doi: 10.1016/j.jcp.2016.05.002
82. Boscheri W. An efficient high order direct ALE ADER finite volume scheme with a posteriori limiting for hydrodynamics and magnetohydrodynamics. *Int J Numer Methods Fluids*. (2017) **84**:76–106. doi: 10.1002/fld.4342
83. Boscheri W, Loubère R. High order accurate direct Arbitrary-Lagrangian-Eulerian ADER-MOOD finite volume schemes for non-conservative hyperbolic systems with stiff source terms. *Commun Comput Phys*. (2017) **21**:271–312. doi: 10.4208/cicp.OA-2015-0024
84. Fambri F, Dumbser M, Zanotti O. Space-time adaptive ADER-DG schemes for dissipative flows: Compressible Navier-Stokes and resistive MHD equations. *Comput Phys Commun*. (2017) **220**:297–318. doi: 10.1016/j.cpc.2017.08.001
85. Tavelli M, Dumbser M, Charrier DE, Rannabauer L, Weinzierl T, Bader M. A simple diffuse interface approach on adaptive Cartesian grids for the linear elastic wave equations with complex topography. *J Comput Phys*. (2019) **386**:158–89. doi: 10.1016/j.jcp.2019.02.004
86. Clain S, Diot S, Loubère R. A high-order finite volume method for systems of conservation laws Multi-dimensional Optimal Order Detection (MOOD). *J Comput Phys*. (2011) **230**:4028–50. doi: 10.1016/j.jcp.2011.02.026
87. Diot S, Clain S, Loubère R. Improved detection criteria for the Multi-dimensional Optimal Order Detection (MOOD) on unstructured meshes with very high-order polynomials. *Comput Fluids*. (2012) **64**:43–63. doi: 10.1016/j.compfluid.2012.05.004
88. Diot S, Loubère R, Clain S. The MOOD method in the three-dimensional case: very-high-order finite volume method for hyperbolic systems. *Int J Numer Methods Fluids*. (2013) **73**:362–92. doi: 10.1002/fld.3804
89. Castro C, Käser M, Toro E. Space–time adaptive numerical methods for geophysical applications. *Philos Trans R Soc A Math Phys Eng Sci*. (2009) **367**:4613–31. doi: 10.1098/rsta.2009.0158
90. Toro EF, Hidalgo A. ADER finite volume schemes for nonlinear reaction-diffusion equations. *Appl Num Math*. (2009) **59**:73–100. doi: 10.1016/j.apnum.2007.12.001
91. Taube A, Dumbser M, Munz CD, Schneider R. A high-order discontinuous Galerkin method with time-accurate local time stepping for the Maxwell equations. *Int J Num Model Electron Netw Devices Fields*. (2009) **22**:77–103. doi: 10.1002/jnm.700

92. Montecinos G, Castro CE, Dumbser M, Toro EF. Comparison of solvers for the generalized Riemann problem for hyperbolic systems with source terms. *J Comput Phys*. (2012) **231**:6472–94. doi: 10.1016/j.jcp.2012.06.011
93. Montecinos GI, Toro EF. Reformulations for general advection-diffusion-reaction equations and locally implicit ADER schemes. *J Comput Phys*. (2014) **275**:415–42. doi: 10.1016/j.jcp.2014.06.018
94. Toro EF, Montecinos GI. Advection-diffusion-reaction equations: hyperbolization and high-order ADER discretizations. *SIAM J Sci Comput*. (2014) **36**:A2423–57. doi: 10.1137/130937469
95. Toro E, Montecinos G. Implicit, semi-analytical solution of the generalized Riemann problem for stiff hyperbolic balance laws. *J Comput Phys*. (2015) **303**:146–72. doi: 10.1016/j.jcp.2015.09.039
96. Toro EF, Castro CE, Lee BJ. A novel numerical flux for the 3D Euler equations with general equation of state. *J Comput Phys*. (2015) **303**:80–94. doi: 10.1016/j.jcp.2015.09.037
97. Busto S. *Contributions to the Numerical Solution of Heterogeneous Fluid Mechanics Models*. Universidade de Santiago de Compostela (2018).
98. Montecinos GI, López-Rios JC, Lecaros R, Ortega JH, Toro EF. An ADER-type scheme for a class of equations arising from the water-wave theory. *Comput Fluids*. (2016) **132**:76–93. doi: 10.1016/j.compfluid.2016.04.012
99. Busto S, Toro EF, Vázquez-Cendón ME. Design and analysis of ADER-type schemes for model advection–diffusion–reaction equations. *J Comput Phys*. (2016) **327**:553–75. doi: 10.1016/j.jcp.2016.09.043
100. Contarino C, Toro EF, Montecinos GI, Borsche R, Kall J. Junction-generalized Riemann problem for stiff hyperbolic balance laws in networks: an implicit solver and ADER schemes. *J Comput Phys*. (2016) **315**:409–33. doi: 10.1016/j.jcp.2016.03.049
101. Busto S, Ferrín JL, Toro EF, Vázquez-Cendón ME. A projection hybrid high order finite volume/finite element method for incompressible turbulent flows. *J Comput Phys*. (2018) **353**:169–92. doi: 10.1016/j.jcp.2017.10.004
102. Dematté R, Titarev VA, Montecinos GI, Toro EF. ADER methods for hyperbolic equations with a time-reconstruction solver for the generalized Riemann problem: the scalar case. *Commun Appl Math Comput*. (2019). doi: 10.1007/s42967-019-00040-x
103. Francois MM, Sun A, King WE, Henson NJ, Tourret D, Bronkhorst CA, et al. Modeling of additive manufacturing processes for metals: challenges and opportunities. *Curr Opin Solid State Mater Sci*. (2017) **21**:198–206. doi: 10.1016/j.cossms.2016.12.001
104. Andreotti B, Forterre Y, Pouliquen O. *Granular Media: Between Fluid and Solid*. Cambridge University Press (2013). Available online at: http://www.edition-sciences.com/milieux-granulaires-entre-fluide-et-solide.htm.
105. Balmforth NJ, Frigaard IA, Ovarlez G. Yielding to stress : recent developments in viscoplastic fluid mechanics. *Annu Rev Fluid Mech*. (2014) **46**:121–46. doi: 10.1146/annurev-fluid-010313-141424
106. Dumbser M. A simple two–phase method for the simulation of complex free surface flows. *Comput Methods Appl Mech Eng*. (2011) **200**:1204–19. doi: 10.1016/j.cma.2010.10.011
107. Gaburro E, Castro M, Dumbser M. A well balanced diffuse interface method for complex nonhydrostatic free surface flows. *Comput Fluids*. (2018) **175**:180–98. doi: 10.1016/j.compfluid.2018.08.013
108. Kemm F, Gaburro E, Thein F, Dumbser M. A simple diffuse interface approach for compressible flows around moving solids of arbitrary shape based on a reduced Baer-Nunziato model. *arXiv [Preprint]*. (2020) arXiv:2001.10326.
109. Gavrilyuk SL, Favrie N, Saurel R. Modelling wave dynamics of compressible elastic materials. *J Comput Phys*. (2008) **227**:2941–69. doi: 10.1016/j.jcp.2007.11.030
110. Favrie N, Gavrilyuk SL, Saurel R. Solid–fluid diffuse interface model in cases of extreme deformations. *J Comput Phys*. (2009) **228**:6037–77. doi: 10.1016/j.jcp.2009.05.015
111. Favrie N, Gavrilyuk SL. Diffuse interface model for compressible fluid - Compressible elastic-plastic solid interaction. *J Comput Phys*. (2012) **231**:2695–723. doi: 10.1016/j.jcp.2011.11.027
112. Ndanou S, Favrie N, Gavrilyuk S. Multi–solid and multi–fluid diffuse interface model: applications to dynamic fracture and fragmentation. *J Comput Phys*. (2015) **295**:523–55. doi: 10.1016/j.jcp.2015.04.024
113. de Brauer A, Iollo A, Milcent T. A cartesian scheme for compressible multimaterial hyperelastic models with plasticity. *Commun Comput Phys*. (2017) **22**:1362–84. doi: 10.4208/cicp.OA-2017-0018
114. Michael L, Nikiforakis N. A multi-physics methodology for the simulation of reactive flow and elastoplastic structural response. *J Comput Phys*. (2018) **367**:1–27. doi: 10.1016/j.jcp.2018.03.037
115. Jackson H, Nikiforakis N. A unified Eulerian framework for multimaterial continuum mechanics. *J Comput Phys*. (2020) **401**:109022. doi: 10.1016/j.jcp.2019.109022
116. Barton PT. An interface-capturing Godunov method for the simulation of compressible solid-fluid problems. *J Comput Phys*. (2019) **390**:25–50. doi: 10.1016/j.jcp.2019.03.044
117. Bungartz HJ, Mehl M, Neckel T, Weinzierl T. The PDE framework Peano applied to fluid dynamics: An efficient implementation of a parallel multiscale fluid dynamics solver on octree-like adaptive Cartesian grids. *Comput Mech*. (2010) **46**:103–14. doi: 10.1007/s00466-009-0436-x
118. Weinzierl T, Mehl M. Peano-A traversal and storage scheme for octree-like adaptive Cartesian multiscale grids. *SIAM J Sci Comput*. (2011) **33**:2732–60. doi: 10.1137/100799071
119. Boscheri W, Dumbser M. Arbitrary–Lagrangian–Eulerian Discontinuous Galerkin schemes with a posteriori subcell finite volume limiting on moving unstructured meshes. *J Comput Phys*. (2017) **346**:449–79. doi: 10.1016/j.jcp.2017.06.022
120. Gaburro E. *Well Balanced Arbitrary-Lagrangian-Eulerian Finite Volume Schemes on Moving Nonconforming Meshes for Non-conservative Hyperbolic Systems*. University of Trento (2018).
121. Jiang G, Shu C. Efficient implementation of weighted ENO schemes. *J Comput Phys*. (1996) **126**:202–28. doi: 10.1006/jcph.1996.0130
122. Balsara D, Shu C. Monotonicity preserving weighted essentially non-oscillatory schemes with increasingly high order of accuracy. *J Comput Phys*. (2000) **160**:405–52. doi: 10.1006/jcph.2000.6443
123. Dumbser M, Käser M. Arbitrary high order non-oscillatory Finite Volume schemes on unstructured meshes for linear hyperbolic systems. *J Comput Phys*. (2007) **221**:693–723. doi: 10.1016/j.jcp.2006.06.043
124. Zhang Y, Shu C. Third order WENO scheme on three dimensional tetrahedral meshes. *Commun Comput Phys*. (2009) **5**:836–48.
125. Shu CW. High order WENO and DG methods for time-dependent convection-dominated PDEs: a brief survey of several recent developments. *J Comput Phys*. (2016) **316**:598–613. doi: 10.1016/j.jcp.2016.04.030
126. Hank S, Gavrilyuk S, Favrie N, Massoni J. Impact simulation by an Eulerian model for interaction of multiple elastic-plastic solids and fluids. *Int J Impact Eng*. (2017) **109**:104–11. doi: 10.1016/j.ijimpeng.2017.06.003
127. Hank S, Favrie N, Massoni J. Modeling hyperelasticity in non-equilibrium multiphase flows. *J Comput Phys*. (2017) **330**:65–91. doi: 10.1016/j.jcp.2016.11.001
128. Stroud A. *Approximate Calculation of Multiple Integrals*. Englewood Cliffs, NJ: Prentice-Hall Inc. (1971).
129. Titarev VA, Tsoutsanis P, Drikakis D. WENO schemes for mixed–element unstructured meshes. *Commun Comput Phys*. (2010) **8**:585–609. doi: 10.4208/cicp.040909.080110a
130. Tsoutsanis P, Titarev VA, Drikakis D. WENO schemes on arbitrary mixed-element unstructured meshes in three space dimensions. *J Comput Phys*. (2011) **230**:1585–601. doi: 10.1016/j.jcp.2010.11.023
131. Levy D, Puppo G, Russo G. Compact central WENO schemes for multidimensional conservation laws. *SIAM J Sci Comput*. (2000) **22**:656–72. doi: 10.1137/S1064827599359461
132. Dumbser M, Boscheri W, Semplice M, Russo G. Central weighted ENO schemes for hyperbolic conservation laws on fixed and moving unstructured meshes. *SIAM J Sci Comput*. (2017) **39**:A2564–91. doi: 10.1137/17M1111036
133. Semplice M, Coco A, Russo G. Adaptive mesh refinement for hyperbolic systems based on third-order compact WENO reconstruction. *J Sci Comput*. (2016) **66**:692–724. doi: 10.1007/s10915-015-0038-z
134. Hu C, Shu C. A high-order WENO finite difference scheme for the equations of ideal magnetohydrodynamics. *J Comput Phys*. (1999) **150**:561–94. doi: 10.1006/jcph.1999.6207
135. Barth TJ, Jespersen DC. *The Design and Application of Upwind Schemes on Unstructured Meshes*. AIAA Paper 89-0366. (1989). p. 1–12.

136. Després B. Polynomials with bounds and numerical approximation. *Numer Algorithms*. (2017) **76**:829–59. doi: 10.1007/s11075-017-0286-0
137. Campos-Pinto M, Charles F, Després B, Herda M. A projection algorithm on the set of polynomials with two bounds. *arXiv preprint* arXiv:190505546 (2019). doi: 10.1007/s11075-019-00872-x
138. Levy D, Puppo G, Russo G. Central WENO schemes for hyperbolic systems of conservation laws. *Math Model Numer Anal.* (1999) **33**:547–71. doi: 10.1051/m2an:1999152
139. Levy D, Puppo G, Russo G. A third order central WENO scheme for 2D conservation laws. *Appl Numer Math*. (2000) **33**:415–21. doi: 10.1016/S0168-9274(99)00108-7
140. Levy D, Puppo G, Russo G. A fourth-order central WENO scheme for multidimensional hyperbolic systems of conservation laws. *SIAM J Sci Comput*. (2002) **24**:480–506. doi: 10.1137/S1064827501385852
141. Cravero I, Puppo G, Semplice M, Visconti G. CWENO: uniformly accurate reconstructions for balance laws. *Math Comput*. (2018) **87**:1689–719. doi: 10.1090/mcom/3273
142. Boscheri W, Semplice M, Dumbser M. Central WENO subcell finite volume limiters for ADER discontinuous Galerkin schemes on fixed and moving unstructured meshes. *Commun Comput Phys*. (2019) **25**:311–46. doi: 10.4208/cicp.OA-2018-0069
143. Jackson H. On the eigenvalues of the ADER-WENO Galerkin predictor. *J Comput Phys*. (2017) **333**:409–13. doi: 10.1016/j.jcp.2016.12.058
144. Banach S. Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fundam Math*. (1922) **3**:133–81. doi: 10.4064/fm-3-1-133-181
145. Dal Maso G, LeFloch PG, Murat F. Definition and weak stability of nonconservative products. *J Math Pures Appl*. (1995) **74**:483–548.
146. Parés C. Numerical methods for nonconservative hyperbolic systems: a theoretical framework. *SIAM J Numer Anal*. (2006) **44**:300–21. doi: 10.1137/050628052
147. Castro MJ, Gallardo JM, Parés C. High-order finite volume schemes based on reconstruction of states for solving hyperbolic systems with nonconservative products. Applications to shallow-water systems. *Math Comput*. (2006) **75**:1103–34. doi: 10.1090/S0025-5718-06-01851-5
148. Dumbser M, Toro EF. A simple extension of the Osher Riemann solver to non-conservative hyperbolic systems. *J Sci Comput*. (2011) **48**:70–88. doi: 10.1007/s10915-010-9400-3
149. Toro EF, Vázquez-Cendón ME. Flux splitting schemes for the Euler equations. *Comput Fluids*. (2012) **70**:1–12. doi: 10.1016/j.compfluid.2012.08.023
150. Hartmann R, Houston P. Adaptive discontinuous Galerkin finite element methods for the compressible Euler equations. *J Comp Phys*. (2002) **183**:508–32. doi: 10.1006/jcph.2002.7206
151. Persson PO, Peraire J. *Sub-cell Shock Capturing for Discontinuous Galerkin Methods*. AIAA Paper 2006-112 (2006).
152. Cesenek J, Feistauer M, Horacek J, Kucera V, Prokopova J. Simulation of compressible viscous flow in time–dependent domains. *Appl Math Comput*. (2013) **219**:7139–50. doi: 10.1016/j.amc.2011.08.077
153. Guermond JL, Nazarov M, Popov B, Tomas I. Second-order invariant domain preserving approximation of the Euler equations using convex limiting. *SIAM J Sci Comput*. (2018) **40**:A3211–39. doi: 10.1137/17M1149961
154. Cockburn B, Shu CW. The Runge-Kutta discontinuous Galerkin method for conservation laws V: multidimensional systems. *J Comput Phys*. (1998) **141**:199–224. doi: 10.1006/jcph.1998.5892
155. Qiu J, Shu C. Runge-Kutta discontinuous Galerkin method using WENO limiters. *SIAM J Sci Comput*. (2005) **26**:907–29. doi: 10.1137/S1064827503425298
156. Qiu J, Shu CW. Hermite WENO schemes and their application as limiters for Runge-Kutta discontinuous Galerkin Method: one-dimensional case. *J Comput Phys*. (2004) **193**:115–35. doi: 10.1016/j.jcp.2003.07.026
157. Balsara D, Altmann C, Munz CD, Dumbser M. A sub-cell based indicator for troubled zones in RKDG schemes and a novel class of hybrid RKDG+HWENO schemes. *J Comput Phys*. (2007) **226**:586–620. doi: 10.1016/j.jcp.2007.04.032
158. Luo H, Baum JD, Löhner R. A hermite WENO-based limiter for discontinuous Galerkin method on unstructured grids. *J Comput Phys*. (2007) **225**:686–713. doi: 10.1016/j.jcp.2006.12.017
159. Krivodonova L. Limiters for high-order discontinuous Galerkin methods. *J Comput Phys*. (2007) **226**:879–96. doi: 10.1016/j.jcp.2007.05.011
160. Zhu J, Qiu J, Shu CW, Dumbser M. Runge-Kutta discontinuous Galerkin method using WENO limiters II: unstructured meshes. *J Comput Phys*. (2008) **227**:4330–53. doi: 10.1016/j.jcp.2007.12.024
161. J Zhu CWS X Zhong, Qiu J. Runge-Kutta discontinuous Galerkin method using a new type of WENO limiters on unstructured meshes. *J Comp Phys*. (2013) **248**:200–20. doi: 10.1016/j.jcp.2013.04.012
162. Boscheri W, Loubère R, Dumbser M. Direct Arbitrary-Lagrangian-Eulerian ADER-MOOD finite volume schemes for multidimensional hyperbolic conservation laws. *J Comput Phys*. (2015) **292**:56–87. doi: 10.1016/j.jcp.2015.03.015
163. Sonntag M, Munz CD. Shock Capturing for discontinuous Galerkin methods using Finite Volume Subcells. In: Fuhrmann J, Ohlberger M, Rohde C, editors. *Finite Volumes for Complex Applications VII*. Cham: Springer (2014). p. 945–53.
164. Rannabauer L, Dumbser M, Bader M. ADER-DG with a-posteriori finite-volume limiting to simulate tsunamis in a parallel adaptive mesh refinement framework. *Comput Fluids*. (2018) **173**:299–306. doi: 10.1016/j.compfluid.2018.01.031
165. de la Rosa JN, Munz CD. Hybrid DG/FV schemes for magnetohydrodynamics and relativistic hydrodynamics. *Comput Phys Commun*. (2018) **222**:113–35. doi: 10.1016/j.cpc.2017.09.026
166. Dumbser M, Peshkov I, Romenski E. A Unified Hyperbolic Formulation for Viscous Fluids and Elastoplastic Solids. In: Klingenberg C, Westdickenberg M, editors. *Theory, Numerics and Applications of Hyperbolic Problems II. HYP 2016. vol. 237 of Springer Proceedings in Mathematics and Statistics*. Springer International Publishing (2018). p. 451–63.
167. Peshkov I, Romenski E, Dumbser M. Continuum mechanics with torsion. *Continuum Mech Thermodyn*. (2019) **31**:1517–41. doi: 10.1007/s00161-019-00770-6
168. Romenski EI. Hyperbolic systems of thermodynamically compatible conservation laws in continuum mechanics. *Math Comput Model*. (1998) **28**:115–30. doi: 10.1016/S0895-7177(98)00159-9
169. Peshkov I, Pavelka M, Romenski E, Grmela M. Continuum mechanics and thermodynamics in the Hamilton and the Godunov-type formulations. *Continuum Mech Thermodyn*. (2018) **30**:1343–78. doi: 10.1007/s00161-018-0621-2
170. Godunov SK. An interesting class of quasilinear systems. *Dokl Akad Nauk SSSR*. (1961) **139**:521–3.
171. Godunov S. Symmetric form of the magnetohydrodynamic equation. *Numer Methods Mech Continuum Medium*. (1972) **3**:26–34.
172. Godunov SK, Romenski EI. Thermodynamics, conservation laws and symmetric forms of differential equations in mechanics of continuous media. *Comput Fluid Dyn Rev*. (1995) **95**:19–31.
173. Godunov SK, Mikhaîlova TY, Romenskiĭ EI. Systems of thermodynamically coordinated laws of conservation invariant under rotations. *Siberian Math J*. (1996) **37**:690–705. doi: 10.1007/BF02104662
174. Romensky EI. Thermodynamics and hyperbolic systems of balance laws in continuum mechanics. In: Toro EF, editor. *Godunov Methods: Theory and Applications*. New York, NY: Springer US (2001). p. 745–61. Available online at: http://www.springer.com/gp/book/9780306466014
175. Hu C, Shu C. Weighted essentially non-oscillatory schemes on triangular meshes. *J Comput Phys*. (1999) **150**:97–127. doi: 10.1006/jcph.1998.6165
176. Sambasivan S, Shashkov M, Burton DE. A finite volume cell-centered Lagrangian hydrodynamics approach for solids in general unstructured grids. *Int J Numer Methods Fluids*. (2013) **72**:770–810.
177. Maire PH, Abgrall R, Breil J, LoubèRe R, Rebourcet B. A nominally second-order cell-centered Lagrangian scheme for simulating elastic-plastic flows on two-dimensional unstructured grids. *J Comput Phys*. (2013) **235**:626–65. doi: 10.1016/j.jcp.2012.10.017
178. Burton DE, Morgan NR, Carney TC, Kenamond MA. Reduction of dissipation in Lagrange cell-centered hydrodynamics (CCH) through corner gradient reconstruction (CGR). *J Comput Phys*. (2015) **299**:229–80. doi: 10.1016/j.jcp.2015.06.041
179. Barton PT, Drikakis D, Romenski EI. An Eulerian finite-volume scheme for large elastoplastic deformations in solids. *Int J Numer Methods Eng*. (2010) **81**:453–84. doi: 10.1002/nme.2695

180. Barton PT, Obadia B, Drikakis D. A conservative level-set based method for compressible solid/fluid problems on fixed grids. *J Comput Phys*. (2011) **230**:7867–90. doi: 10.1016/j.jcp.2011.07.008
181. Dobrev VA, Kolev TV, Rieben RN. High order curvilinear finite elements for elastic–plastic Lagrangian dynamics. *J Comput Phys*. (2014) **257**:1062–80. doi: 10.1016/j.jcp.2013.01.015
182. Ferrari A, Dumbser M, Toro EF, Armanini A. A new 3D parallel SPH scheme for free surface flows. *Comput Fluids*. (2009) **38**:1203–17. doi: 10.1016/j.compfluid.2008.11.012
183. Jackson H, Nikiforakis N. A numerical scheme for non-Newtonian fluids and plastic solids under the GPR model. *J Comput Phys*. (2019) **387**:410–29. doi: 10.1016/j.jcp.2019.02.025
184. Schmidmayer K, Petitpas F, Daniel E, Favrie N, Gavrilyuk S. Iterated upwind schemes for gas dynamics. *J Comput Phys*. (2017) **334**:468–96. doi: 10.1016/j.jcp.2017.01.001
185. Chiocchetti S, Peshkov I, Gavrilyuk S, Dumbser M. High order ADER schemes and GLM curl cleaning for a first order hyperbolic formulation of compressible flow with surface tension. *arXiv [Preprint]*. (2020). Available online at: https://arxiv.org/abs/2002.08818
186. Romenski E, Resnyansky AD, Toro EF. Conservative hyperbolic formulation for compressible two-phase flow with different phase pressures and temperatures. *Q Appl Math*. (2007) **65**:259–79. doi: 10.1090/S0033-569X-07-01051-2
187. Romenski E, Drikakis D, Toro EF. Conservative models and numerical methods for compressible two-phase flow. *J Sci Comput*. (2010) **42**:68–95. doi: 10.1007/s10915-009-9316-y
188. Abbate E, Iollo A, Puppo G. An asymptotic-preserving all-speed scheme for fluid dynamics and nonlinear elasticity. *SIAM J Sci Comput*. (2019) 41:A2850–79. doi: 10.1137/18M1232954
189. Tavelli M, Dumbser M. A pressure-based semi-implicit space-time discontinuous Galerkin method on staggered unstructured meshes for the solution of the compressible Navier-Stokes equations at all Mach numbers. *J Comput Phys*. (2017) **341**:341–76. doi: 10.1016/j.jcp.2017.03.030
190. Ioriatti M, Dumbser M. A posteriori sub-cell finite volume limiting of staggered semi-implicit discontinuous Galerkin schemes for the shallow water equations. *Appl Numer Math*. (2019) **135**:443–80. doi: 10.1016/j.apnum.2018.08.018
191. Busto S, Tavelli M, Boscheri W, Dumbser M. Efficient high order accurate staggered semi-implicit discontinuous Galerkin methods for natural convection problems. *Comput Fluids*. (2020) **198**:104399. doi: 10.1016/j.compfluid.2019.104399

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

MEDICAL PHYSICS AND IMAGING

Dr. Roberta Frass-Kriegl studied Engineering Physics at Vienna University of Technology (BSc 2009) and Technical University of Munich (MSc 2011), and holds a PhD degree in Medical Physics (2014) jointly awarded by the Medical University of Vienna, Austria, and University Paris-South, France. She is postdoctoral researcher at the Medical University of Vienna specialized in MRI hardware and software development, and initiated the MR Simulation Lab in 2018.

Bernhard Gruber is a PhD student in Medical Physics at the Medical University Vienna and holds an MSc (2016) in Medical Engineering (Mechatronics). Since 2013, Bernhard was involved in several projects at the JKU Linz (Austria), the UMC Utrecht (The Netherlands) and the MGH Martinos Center (USA) - also in cooperation with several industrial partners. He is contributor to several open source projects in the MRI community.

Dr. Elmar Laistler holds an MSc (2005) and PhD (2011) in Physics from Vienna University of Technology, Austria, is the head of the RF Lab at the Medical University of Vienna and has been guest researcher in Paris (France) and Berlin (Germany). His research focus lies in hardware and software development for UHF MR systems, including electromagnetic simulation.

Lena Nohava received her Physics diploma from the University of Vienna in 2017 and, currently, is a PhD student in the BioMaps laboratory at the University Paris-Saclay, France, working in collaboration with the RF Lab at the Medical University of Vienna. Her main research interests include UHF MR hardware, especially RF coils, and electromagnetic simulations.

Dr. Sigrun Roat received a Master's degree in Mathematics (2011) from Vienna University of Technology and a PhD in Medical Physics (2015) from the Medical University of Vienna, Austria, combining her mathematical training with hands-on physics in a medical environment. Her research focus lies in the development of multi-nuclear radio frequency coils for UHF MR spectroscopy, including full wave numerical optimization and validation.

Dr. Mathieu Sarracanie received his PhD in Physics 2011 from the Université Paris-Sud, France, followed by a post-doctoral position at Harvard University, Department of Physics and MGH/HST A. A. Martinos Center for Biomedical Imaging, from 2011 to 2016. He was recently appointed assistant professor and is currently leading the AMT Lab jointly with Prof. N. Salameh, Dept. of Biomedical Engineering, University of Basel (Switzerland), with a research focus on adaptable MRI technology for medical diagnosis.

Dr. Peter Hömmen holds a Master of Science degree in Biomedical Engineering (2015) from the Ilmenau University of Technology, Ilmenau, Germany. Currently he is a research associate working on "Current density imaging with SQUID-based ultra-low field MRI" at the Physikalisch-Technische Bundesanstalt, Berlin, Germany, and pursues a doctorate at the Ilmenau University of Technology.

Antti J. Mäkinen holds a Master of Science degree in Biomedical Engineering (2015) from the Aalto University School of Science, Finland. During a research exchange at the Physikalisch-Technische Bundesanstalt, Berlin, Germany, in 2018 he started collaborating on current density imaging. Currently he is a doctoral candidate at the Departments of Neuroscience and Biomedical Engineering, MEG–MRI research group, Espoo, Aalto University, Finland.

Roberta Frass-Kriegl is working as PostDoc researcher at the Center for Medical Physics and Biomedical Engineering at the Medical University of Vienna, Austria. Specialized on hardware and software development for magnetic resonance imaging, she initiated the MR Simulation Lab in 2018. She studied Engineering Physics at Technische Universität Wien (BSc 2009), Austria, and Technische Universität München (MSc 2011), Germany, and holds a PhD degree in Medical Physics (2014) jointly awarded by the Medical University of Vienna, Austria, and Université Paris-Sud, France.

Sigrun Roat received a Master's degree in Mathematics (2011) from Vienna University of Technology, Austria, and a PhD in Medical Physics (2015) from the Medical University of Vienna, Austria, combining her mathematical training with hands-on physics in a medical environment. Her research focus lies in the development of multi-nuclear radio frequency coils for UHF MR spectroscopy, including full wave numerical optimization and validation.

Bernhard Gruber is a PhD student in Medical Physics at the Medical University Vienna, Austria, and holds an MSc (2016) in Medical Engineering (Mechatronics). Since 2013, Bernhard was involved in several projects at the JKU Linz, Austria, the UMC Utrecht, The Netherlands, and the MGH Martinos Center, Boston, USA - also in cooperation with several industrial partners. He is a contributor to several open source projects in the MRI community.

Lena Nohava received her Physics diploma from the University of Vienna, Austria, in 2017 and is now a PhD student in the BioMaps laboratory at the University Paris-Saclay, France, working in collaboration with the RF Lab at the Medical University of Vienna, Austria. Her main research interests include UHF MR hardware, especially RF coils, and electromagnetic simulations.

Elmar Laistler holds an MSc (2005) and PhD (2011) in Physics from Vienna University of Technology, Austria, and is currently the head of the RF Lab at the Medical University of Vienna, Austria. Recently, he has been a guest researcher in Paris-Saclay and at PTB Berlin. His research focus lies in hardware and software development for UHF MR systems, including electromagnetic simulation.

Jean-Christophe Ginefri received his Ph.D. in physics from the Université Paris Sud, Orsay (France) in 1999. He is associate professor at Paris-Saclay University and his research activities are in the field of instrumental and methodological development for MRI, more specifically focused on the development of highly sensitive NMR detection systems.

Peter Hömmen holds a Master of Science degree in Biomedical Engineering from the Ilmenau University of Technology, Ilmenau, Germany. Currently he is a research associate working on "Current density imaging with SQUID-based ultra-low field MRI" at the Physikalisch-Technische Bundesanstalt, Berlin, Germany, and pursues a doctorate at the Ilmenau University of Technology.

Antti J. Mäkinen holds a Master of Science degree in Biomedical Engineering from the Aalto University School of Science, Finland. During a research exchange at the Physikalisch-Technische Bundesanstalt, Berlin, Germany, in 2018 he started collaborating on current density imaging. Currently he is a doctoral candidate at the Departments of Neuroscience and Biomedical Engineering, MEG–MRI research group, Espoo, Aalto University, Finland.

Mathieu Sarracanie received his PhD in Physics 2011 from the Université Paris-Sud, France, followed by a post-doctoral position at Harvard University (Dept. of Physics) and MGH/HST Athinoula A. Martinos Center for Biomedical Imaging (USA) from 2011 to 2016. He was recently appointed assistant professor and is currently leading the AMT Lab jointly with Prof. N. Salameh, Dept. of Biomedical Engineering, University of Basel (Switzerland), with a research focus on adaptable MRI technology for medical diagnosis."

Due to the inherently low sensitivity of the technique and the fast development of superconducting magnet technology over the past 40 years most clinically used systems now operate between 0.2 Tesla (T) and 7 T. Speed and brilliance of images are important features in clinical MRI, which has led to a race to ever higher B 0. However, there has been a long standing debate on what would be the optimum B 0 in terms of (clinical) image contrast. Therefore, a vivid international research community is working to improve ultra-low field (ULF) MRI, reviewed elegantly by

Sarracanie & Salameh describe most recent developments at low and ultra-low field, covering a broad spectrum from pre-polarized MRI, ultra-sensitive sensors, and their relevance in clinical application. This work could result in cost-effective alternatives to high-and ultrahigh-field (UHF) MRI, opening new perspectives through novel image contrast to complement established diagnostic tools.

Hömmen & Mäkinen et al. in this issue even evaluate the performance of zero-field-encoded ULF-MRI for in vivo 3D current density imaging and potentially conductivity mapping of the human head. Although their research reveals that image artifacts may affect reconstruction quality, their simulations also indicate that current-density reconstruction in the scalp requires spatial resolution less than 5 mm and demonstrate that the necessary sensitivity coverage could already be accomplished by multi-channel devices today.

At the most frequently used clinical field strength of 1.5 T, sensitivity and speed can be further improved via radio-frequency (RF) coil arrays. Gruber et al. attempt to solve a practical problem in clinical imaging, namely the varying size of patients, by proposing a size-adaptive, one-size-fits-all flexible coil array for knee MRI. Built of partly-stretchable loops, the novel array coil shows an improved SNR of up to 100 % in 20 mm depth from the phantom surface, demonstrating the effectiveness of adaptive RF coils by reducing noise contributions from empty coil regions. Although the use of array coils has been extremely successful, there are technological challenges at higher B 0, in particular regarding the integration of a large number of coil elements. Cabling including bulky cable traps renders such large coil arrays rather heavy and rigid. Frass-Kriegl et al. present a novel concept for reducing the number of coil elements dubbed "multi-loop coils" (MLC), exploiting the high sensitivity of small RF coils while reducing sample induced noise together with an extended field of view. Investigations were performed using MLCs, each composed of multiple smaller loops, targeting MRI at high (3 T) and at ultra-high field strength (7 T). Such MLCs appear advantageous for the development of single RF coils but also individual elements of arrays, especially for applications with a larger area and shallow target depth, such as skin imaging or high-resolution MRI of brain slices.

Nohava et al. review the problem of how to get rid of the numerous cables and cable traps completely by transmitting the MR signal wirelessly from the coil. Such RF coil arrays have the potential to be much lighter and easier in terms of handling in a clinical environment. The paper addresses the scientific and technological challenges of wireless RF coils for MRI, including the MR receive signal chain, control signaling, and on-coil power supply. They conclude that completely wireless RF coils will ultimately become feasible, however, innovations are required specifically regarding wireless communication technology, MR compatibility, and wireless power supply.

The studies described so far have been performed at the proton resonance frequency as it provides the highest sensitivity for MRI. At UHF, however, multinuclear MRI and MRS becomes feasible. Roat et al. compare a wide range of potential array designs to build a flexible array for cardiac 31 P MR spectroscopy at 7 T, to be integrated in an existing 12-channel proton array. This is an excellent example of combining interdisciplinary skills and one of the most thorough papers on UHF RF hardware development published so far.

In summary, the combination of young talent, interdisciplinary and international research collaboration provides novel and innovative approaches to tackle a broad range of challenges humankind is facing now and in the near future.

frontiers in Physics

REVIEW
published: 12 June 2020
doi: 10.3389/fphy.2020.00172

Low-Field MRI: How Low Can We Go? A Fresh View on an Old Debate

Mathieu Sarracanie and Najat Salameh*

Center for Adaptable MRI Technology, Department of Biomedical Engineering, University of Basel, Allschwil, Switzerland

For about 30 years, MRI set cruising speed at 1.5 T of magnetic field, with a gentle transition toward 3 T systems. In its first 10 years of existence, there was an open debate on the question of most relevant MRI field strengths considering the gain in T_1 contrast, simpler cooling strategies, lower predisposition to generating image artifacts, and naturally cost reduction of small footprint low field systems. At the time, the inherent gain in sensitivity of high field, which would translate in more signal per unit time, quickly ended this debate. The promise of rapid exams or higher image resolution within a reasonable time won over other considerations and set the standards for MR value. Yet, many reasons bring low field MRI in a situation quite different from 40 years ago. From the achieved progress regarding all aspects of MRI technology, an MR scan at 1.5 T in the mid 1980s has very little in common with the equivalent scan in 2020. That clearly indicates that field strength alone is not what drives performance. It is also unlikely that the total number of machines worldwide will grow so to follow the increasing demand considering their overall cost (∼$1M/T). The natural trend is to better control medical expenses worldwide, and reconsidering low-field MRI could lead to the democratization of dedicated, point-of-care devices to decongest high-field clinical scanners. In the present article, we aim to draw an extensive portrait of most recent MRI developments at low (1–199 mT) and ultra-low field (micro-Tesla range) outside of the commercial sphere, and we propose to discuss their potential relevance in future clinical applications. We will cover a broad spectrum from pre-polarized MRI using ultra-sensitive magnetic sensors up to permanent and resistive magnets in compact designs.

Keywords: MRI, low magnetic field, ultra-low field MRI, MR value, point-of-care MRI

OPEN ACCESS

Edited by:
Ewald Moser,
Medical University of Vienna, Austria

Reviewed by:
Angelo Galante,
University of L'Aquila, Italy
Marie Poirier-Quinot,
Université Paris-Sud, France

****Correspondence:***
Mathieu Sarracanie
mathieu.sarracanie@unibas.ch

Specialty section:
This article was submitted to Medical Physics and Imaging, a section of the journal Frontiers in Physics

Received: *04 February 2020*
Accepted: *23 April 2020*
Published: *12 June 2020*

Citation:
Sarracanie M and Salameh N (2020) Low-Field MRI: How Low Can We Go? A Fresh View on an Old Debate. Front. Phys. 8:172. doi: 10.3389/fphy.2020.00172

INTRODUCTION

Low field MRI? The rationale behind it has been around for quite some time, traditionally perceived as a mean to reduce cost or to provide open access to patients suffering from claustrophobia. Over the past 30 years, scientists have supported low-field MRI on multiple occasions and brought facts that corroborate clinical relevance [1–4]. Yet, low field MRI has not spread. Reasons that have been invoked are diverse and have led to numerous debates. From the manufacturer point-of-view, the current business model in MRI results in higher margins allowing to increase profit [5]. From a clinical and academic point-of-view, the quest for higher and higher spatial resolution has led radiologists and scientists worldwide to always push toward high- and ultra-high field MRI research, eventually dominating over all others in peer reviewed journals [6]. One sure thing is that the statistics of MRI sales over the last two decades have certainly helped closing that debate. Nowadays high-field MRI sales ($B_0 \geq 1.5$ T) represent about 85% of the market

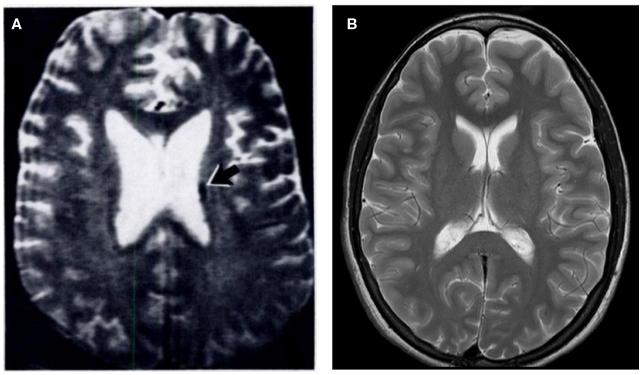


FIGURE 1 | T_2-weighted MR brain images acquired at 1.5 T in **(A)** 1986. Image reused, with permission, from Zimmerman et al. [8] and **(B)** 2009 (authors' database).

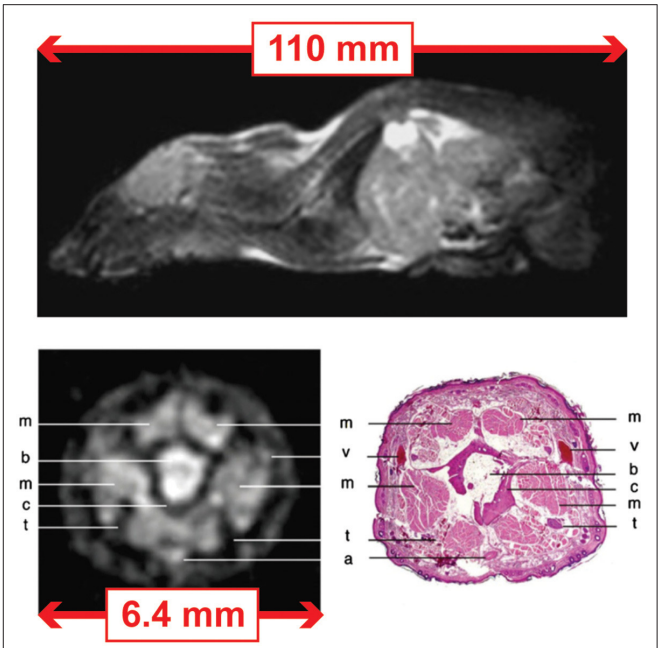


FIGURE 2 | MRI images of the mouse acquired at 0.1 T using a FISP sequence and dedicated coils for the whole-body and tail. (Top) Whole body: field-of-view (FoV) of 110 mm and in plane resolution of 430 × 430 μm^2. The acquisition time was 30 min for 30 slices of ∼1 mm thickness. (Bottom) Tail: field-of-view (FoV) of 6.4 mm and in plane resolution of 100 × 100 μm^2. The acquisition time was 1 h 30 min for 26 slices of ∼ 750 μm thickness. Images modified, with permission, from Choquet et al. [9].

size in Europe and North America [7]. One of the main misconceptions is that low-field MRI translates into poor image resolution, often associated with poor image quality. It is important, as scientists, to state that this concept is purely and simply wrong. Magnetic field strength has by no means ever been a limit to an achievable image resolution. A brief jump into the early days of MRI is enough to appreciate the tremendous leap in image quality that was made for a given field strength (**Figure 1**). More recent work from Choquet et al. [9] has reported MRI of different mouse body parts *in vivo* at 0.1 T (∼4.3 MHz) with down to 100 × 100 × 750 μm^3 voxel size, more than 10 years ago (**Figure 2**). Sensitivity though, and how far the signal lies above the detection chain's noise floor will tell about one's capability to achieve a given resolution in the minimum amount of time. Hence time really is the argument at stake when considering lower field options. Indeed, lower field strengths result in lower bulk magnetization of nuclear spins leading in turn to a reduced sensitivity. Assuming a fixed noise floor in the detection chain, the decreased total magnetic moment brings the maximum signal detectable closer to the latter and the overall signal-to-noise ratio (SNR) drops. One main alternative to compensate for this loss is signal averaging. It is generally accepted that n averages will produce an SNR gain of $\sqrt{n}$. Hence time is currently the true limit to a wide spread of low-field MRI, due to lower overall NMR sensitivity. However, this is also a matter of perspective. Why time considerations have become key in clinical diagnosis has to be contextualized in the current landscape of MRI. Most hospitals currently host one or two MRI scanners, which cost roughly scales with magnetic field (∼$1M/T). As such MRI units are expensive and used for the imaging of all body parts, they are likely to represent a bottleneck in clinical workflows. Hence, the time needed for one scan has to be short in order to scan as many patients as possible within a day. Nowadays, no one can afford a machine that would perform slower than the state-of-the art because there is such a high demand for non-invasive radiation-free diagnosis. Yet, other than applications where speed is truly paramount such as for cardiovascular applications, or patients with a life-threatening risk, fast imaging is only required due to scanners being a low-volume/high price equipment. One could argue that this quest for speed is not as relevant if numerous low-field, low-cost devices were to be used (high-volume, low-price). After all, if the price of a scanner is divided by two and the acquisition time multiplied by two, then the cost per unit time stays the same, and the same number of patients can be scanned within the same time slot. Only cost for personnel would increase. The situation in China is a good illustration of this point: the high population density requires a higher density of MR units, and mid-field MR units represent about 50% of the market size [10], vs. 6% in Europe and North America [7]. As a consequence, depending on the market ability to embrace such a change in paradigm, that approach would naturally reduce pressure on acquisition times. Most importantly, rather than acquisition time or image resolution, one key aspect in the democratization of low-field MRI resides in its value. Most likely, if manufacturers and end users would foresee higher value in low-field MRI solutions, there would be a straightforward path toward mass adoption. Value though is a complex concept that finds different resonances across populations and cultures. One should agree on a simple description of value as being the ratio of a benefit over a cost [11]. For low-field technology to be truly visible and adopted, its value in MRI diagnosis would hence need to be increased. Two approaches then allow to increase value; 1-reduce cost, 2-increase benefits, or both at the same time. Considering cost, the last two decades have already shown

that relevant diagnosis can be achieved in lower-cost lower-field devices [12–17]. Interest though, is hard to trigger if value increases only slightly and yet no broad adoption has followed since. For example, commercially available low-field scanners rely on permanent magnet technology that can weigh up to 13 tons. Thus, their individual cost (that includes siting) has never reached a point where value goes through the roof and triggers such a cultural change. Maybe the economic pressure on health expenses will change the current landscape with populations worldwide aging and growing, but this has been a long-heard argument never followed by action. Eventually, it appears challenging to spark interest in both radiologists and academics only by lowering cost as this is often perceived as leading to less potent technology, except maybe when the research is directed toward developing countries. The latter field of research is considered a niche though, and if cost is one key to selling in these countries, "low-cost" alone will never replace tailor-made solutions to country-specific needs. The alternative to increased value is then to increase benefits. Nowadays, MRI units require specific siting from their heavy weight and intense magnetic field strength, and shielding from magnetic and electromagnetic disturbances. MRI systems are known to be incompatible with most devices unless they are specifically made MRI-compatible. Increasing accessibility by means of a (much) lower footprint, little siting requirement, or enhanced compatibility certainly is a path toward increased benefits, and hence increased value. Now, what key element is driving such heavy weights, compatibility aspects, and ultimately the cost of MRI machines nowadays? Magnetic field is. As a result, low magnetic field MRI could very likely bring high value from both decreased costs and enhanced benefits. But how low can we go? In this manuscript, we aim to provide a fresh view on this old debate in MRI.

SNR, THE ELEPHANT IN THE ROOM

SNR indeed, is the elephant in the room when it comes to low-field MRI. Lower Boltzmann distribution infers lower induced voltages in inductive detection that translates into lower signal. A generally accepted assumption is that SNR scales with the static magnetic field $B_0^{\frac{7}{4}}$ for frequencies above 5 MHz, and $B_0^{\frac{3}{2}}$ at low frequencies (below ∼5 MHz) [18]. One of the main challenges for low-field MR experts is thus to compensate for the loss in SNR per unit time inherited from the reduced magnetic field. Recent work laid an exhaustive portrait of mid-field MRI in the range 0.25–1 T [19], pointing out the current gain of interest for alternative solutions. Increasing SNR via magnets that are not "too" weak certainly is an approach to preserve signal. The latter has some merit considering the tremendous amount of progress magnetic resonance has benefited from over the last three decades, as illustrated in **Figure 1** and in [15]. As a result, most recent developments will translate easily in mid-field regimes which detection physics and design considerations are close to ^{1}H NMR at most widely spread 1.5 T. That said, the big difference in images obtained between the 80's and today at 1.5 T also highlights how field strength is not a guarantee for good image quality, and how much it has to be balanced with high performance acquisition techniques, high sensitivity detectors, image processing, and modern electronics all combined together. Another comment on medium range 0.25–1 T MRI is that it might not allow to move sufficiently far from typical constraints found in high-field MRI. Among them, exclusive magnets made of superconductors for fields higher than 0.5 T [19], magnet weight, low tolerance to magnetic environments and magnetic susceptibility effects. One should keep in mind that an interesting approach to tame SNR is favoring regimes sufficiently different from inductive detection passed 15 MHz, with a focus on noise considerations. As an example, noise predominance below 5 MHz for a variety of coil sizes comes from the coil when body noise predominance is the general rule in mainstream high-field MRI [20]. Current acceleration methods for clinical MR imaging such as parallel imaging are even dictated by sample noise predominance as each coil sees coherent noise from the sample. In the present manuscript, in order to complement existing review work and report on the latest and most active low and ultra-low field MRI research, the authors have narrowed down the span of low-field MRI research considered to articles published within the past 5 years at field < 0.2 T (< 8.5 MHz). More specifically, the authors are reporting on work that already have produced images (even if not yet clinically relevant or only in phantoms), yet excluding simulated work.

HARDWARE CONSIDERATIONS

Magnets

The MRI community shows a growing interest for small footprint MRI technology leveraging low-magnetic fields. Worldwide, an effort has started to spread from hardware and engineering considerations, notably regarding magnet construction.

Permanent Magnets

One of the main clinical application envisioned for small footprint, point-of-care MRI is neuroimaging. Many academic sites driving this effort have opted for permanent magnet architectures, for the most part deriving designs from recent work making use of Halbach geometries [21]. Permanent magnets are of particular interest because they do not require power to produce the spin-polarizing static magnetic field B_0. Zimmerman-Cooley et al. [21] have reported on two generations of lightweight magnets. The first generation uses the intrinsic inhomogeneity found in Halbach magnet geometries for spatial encoding, whereas the second version features a constant 1D spatial encoding magnetic field gradient (SEM) superimposed on the static magnetic field B_0 [22]. In this second iteration, 2D encoding is obtained either by physically rotating the native 1D SEM around the object of interest, or by using a custom made gradient insert avoiding physical rotation [22]. Three-dimensional encoding is obtained from the same gradient insert, from an additional coil arrangement in the last spatial direction. The magnet features a 29-cm diameter bore and weighs 122 kg for an average B_0 of 79.3 mT (3.4 MHz). Field homogeneity is 27,800 ppm (∼95 kHz) over a 20-cm diameter spherical volume (DSV). Imaging capability is still in its infancy but several promising

approaches are being assessed. Recent work from O'Reilly et al. [23] similarly report on a 27-cm diameter bore Halbach array magnet construction. The latter proposes a classic approach to magnet design used at high-field, where the static magnetic field homogeneity is being optimized first, and complemented by a gradient insert featuring coils for spatial encoding in 3 dimensions. The presented magnet weighs ∼ 75 kg for an average magnetic field of 50.4 mT (2.15 MHz). The measured magnetic field shows ∼2,500 ppm (∼5 kHz) homogeneity across a 20-cm DSV. Pushing toward more compact designs, McDaniel et al. [24] have designed and built a head-size, hemispheric magnet consisting of an assembly of NdFeB blocks inserted into a 3D printed former. The magnet weighs 6.3 kg with a mean B_0 of 63.6 mT (2.67 MHz). Over the targeted ROI of $\sim 3 \times 8 \times 8$ cm^3, B_0 covers a 69,200 ppm (∼190 kHz) range. The magnet was designed with a built-in field gradient of ∼117 mT/m. For the remaining spatial encoding directions, two single-sided, hemispherically-shaped gradient coils were wound on the outside of the magnet former. Further away from applications *in vivo*, yet with a major emphasis on portability and democratization of MR technology, Greer et al. [25] have presented a hand-held MR system for 2D imaging (projection) in ∼ 9-mm diameter objects using permanent magnets arranged in a curved single-sided geometry. This device follows the steps of the previously released NMR-MOUSE [26, 27]. After numerical optimization, five NdFeB magnets are arranged so that the sample of interest is partially enclosed by the magnet, allowing to obtain a higher B_0 of 186 mT (8 MHz). The 2D field of view is 9×9 mm^2 with ∼58,500 ppm homogeneity (∼460 kHz) over a 5-mm depth ($G_0 = 2,200$ mT/m). The system features two planar gradients coils printed on PCB. The typical pitfalls for permanent magnet constructions rely in the magnetic field being constantly turned-on, weight, important field inhomogeneity, and poor temperature stability. Weight can be mitigated in purpose-built small size scanner, yet temperature can yield several kilohertz frequency shifts per degree caused both by environmental changes and heating from other components of the scanner, such as gradient coils. Strategies to regulate magnet temperature or to account for resulting frequency drifts in imaging or post-processing pipelines will be crucially needed in order for these technologies to be democratized.

Electromagnets

Electromagnets are an equally relevant alternative for low-field MRI, generally providing better field homogeneity than permanent magnets. In terms of flexibility, they can be turned on and off at will, and the generated field can be modulated by varying the input current in the magnet coils. Electromagnets operating at low field are particularly interesting as requirements on power and cooling can be drastically reduced when their physical footprint decreases. In recent work, Lother et al. [28] report on low field MRI for neonatal applications by means of a bi-planar electromagnet geometry mounted on a steel housing. Planar gradient coils are embedded for spatial encoding in three dimensions over a 140-mm field of view (FoV). The magnet homogeneity over the desired FoV was measured to be 1, 200 ppm (1,100 Hz) for a static magnetic field of 23 mT (965 kHz). The total weight of the magnet with accompanying electronics, acquisition console, and table trolley is below 300 kg. Sarracanie et al. [29] have reported on ultra-low field MRI at 6.5 mT (276 kHz) in a resistive magnet. Originally designed for hyperpolarized ^{3}He MRI in order to assess posture dependence of lung ventilation [30], the geometry was set such that an adult human being could stand inside the magnet. The magnet features a bi-planar geometry with two pairs of ring coils facing each other. The outer ring coils have a diameter of ∼2 m, with 79-cm inter-coil separation. In its original design, each side including coil and flange weighed ∼340 kg. In its final version, the magnet is fully open and quite compact compared to standard clinical scanners. The magnet homogeneity was measured to be 350 ppm (96 Hz) over a 25-cm DSV. In 2015, Galante et al. [31] demonstrated proof-of-concept of very-low field MRI with a scaled down electromagnet compatible with magneto-encephalography (MEG). Their electromagnet is a 23.4-cm diameter wound-solenoid providing $B_0 = 8.9$ mT (373 kHz) in its center. The measured B_0 homogeneity inside a 6-cm region of interest was ∼150 ppm (57 Hz). The X-Y gradient coils are located on the inner surface of the solenoid. The Z gradient is a compensated Maxwell pair configuration placed outside the main coil. With enhanced flexibility and field homogeneity that can be key to imaging performance, electromagnet main weaknesses lie in their need for power supplies that can quickly reach three-phase power and the necessity for liquid cooling, of course depending on the geometry and the field strength envisioned. Yet, rather simple water-cooling can be used that remains more advantageous than complex and expensive cryogenics.

Pre-polarizing Magnets

MRI in the μT range and developments for MEG compatible systems most commonly rely on pre-polarization strategies. In general, electromagnets are used for pre-polarization but experimental setups with permanent magnet also exist where the sample is shuttled from the magnet to the imaging site. Low effort is required on the intrinsic magnetic field homogeneity of such magnets as they only serve to boost spin magnetization before acquisition, thus simplifying their design and production. Pre-polarizing magnets for ULF MRI can be found to operate in a variety of field strengths, from 11 mT to 2 T [32–41], with cooling strategies observed from 20 mT and above.

Low-Frequency Detection

With sensitivity going down from lower spin polarization, detection is one of the key elements to consider that affects imaging performance, from sensor design to signal amplification and noise reduction strategies.

MRI <1 mT

One aspect of ULF NMR and MRI in the μT range focuses on the use of high-sensitivity magnetic sensors to compensate for the extremely weak nuclear spin polarization. To date, the most advanced work that produced images in humans *in vivo* was using SQUIDs [36, 38, 40, 42–44]. Most recent work includes imaging from a seven-channel low-T_c SQUID based system [38] *in vivo* in the human brain at 200 μT (8 kHz)

in a magnetically shielded room (MSR). The SQUID sensors are commercial CE2Blue (Supracon AG, Jena, Germany) with second-order axial gradiometers including a ∼90-mm pick-up loop diameter and baseline. In a very recent paper, Hömmen et al. [45] have used a two-stage low-T_c SQUID sensor at ∼39 μT (1,645 Hz), consisting of a single front-end SQUID with double-transformer coupling read out by a 16-SQUID array [46]. As reported by the authors, the SQUID is housed inside a niobium capsule to shield it from the high polarizing fields. The integrated input coil is connected to a 2nd-order axial gradiometer with 45-mm diameter and 120-mm overall baseline. Oyama et al. [47] have reported on ULF MRI in rat heads that is compatible with MEG. The sensor is a low-T_c dc SQUID with a second-order axial-type gradiometer 15-mm diameter pickup coil installed in a cryostat. Work from Kawagoe et al. [48] reports on the use of a high temperature superconductor (HTS) SQUID combined with an LC resonator to extend their detection area (by∼ ×1.5). Their resonator is composed of a coil and a capacitor set at 1,890 Hz resonance frequency. The signal is detected by the coil inductively coupled to the HTS SQUID and immersed in liquid nitrogen. 2D imaging is performed in a 35-mm diameter water phantom, while the entire setup sits either in an MSR or a compact magnetically shielded box offering ×1.3 more SNR. Another very recent study aims at addressing short relaxation times in oil for food contaminant inspection using an HTS SQUID system equipped with a non-resonant flux transformer [49]. In the latter case, the MR signal from a sample is detected by a pickup coil and transferred to a separately located SQUID via a superconducting input coil. With an attempt to move away from cumbersome MSRs, Liu et al. [50] report on the use of a static magnetic gradient tensor detection and compensation system to stabilize temporal magnetic field fluctuations. With this compensation, ULF MRI could be demonstrated in a 38-mm phantom. Acquisition was performed by a low-T_c hand-wound second-order axial gradiometer inductively coupled to a dc SQUID. The gradiometer was located at the bottom of a fiberglass cryostat, immersed in liquid helium. Parallel efforts in the μT range have kept pre-polarizing strategies to boost nuclear spin polarization, however transitioning toward simpler technology such as atomic magnetometers or even inductive detection to branch out from costly and impractical requirement such as cryogenics. When this community was still very active, images *in vivo* were produced in the human hand and head [51–54]. Recent work from 2017 reports on optically pumped atomic magnetometers (OPAM) combined with liquid cooled pre-polarization coils [55]. Two MSRs host the OPAM and the MRI systems separately. The OPAM sensor uses two laser beams, respectively, the pump and probe lasers. It relies on the magneto-optical effect which leads to the rotation of the linearly polarized plane of the probe laser by an angle proportional to the magnetic field it experiences. The OPAM and the MRI systems are electrically connected by a flux transformer consisting of a second order gradiometer as input coil with a baseline of 175 mm using solenoid coils. The OPAM module operates at 117 μT (5 kHz) by applying a bias field of ∼25 μT (∼1,080 Hz). With respect to inductive detection, pre-polarized μT MR finds application in the industry where it can be used to monitor water fouling [33]. In 2015, Benli et al. [32] had used the same commercial system (Terranova, Magritek, Wellington, New Zealand) for MRI at their local earth magnetic field (47 μT or ∼2 kHz) for teaching purpose [32]. In both work, inductive detection is made from a single channel 84-mm diameter solenoid tuned and matched at ∼2 kHz.

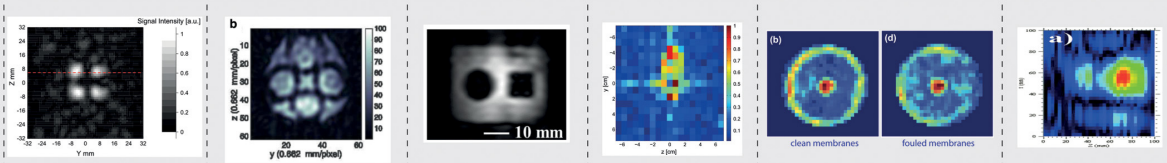
MRI From 1 mT to 199 mT

In all of most recent reported work, inductive detection was chosen with a variety of approaches typical of higher field MR research: separated transmit and receive, transceiver coils, surface, or volume geometries. Diverse designs such as multi-turn loop coils [25], saddle coils [56], multiple-channel phased-arrays [21], solenoids [23, 28], or custom spiral volume coils [24, 29] were employed. In 2015, Zimmerman-Cooley and colleagues built a 25-turn, 20-cm diameter, and 25-cm long solenoid coil for transmit, and a 14-cm diameter multi-channel receive array made of 8 8-cm diameter, overlapping loops [21]. The coils were tuned and matched at 3.29 MHz. Later, Stockmann introduced the idea of combining swept WURST RF pulse echo trains for Transmit Array Spatial Encoding (TRASE) in very inhomogeneous B_0 fields, and was able to demonstrate the acquisition of spatially encoded 1D profiles [57]. In 2015, Sarracanie and colleagues designed an innovative single channel, head-shaped, spiral volume transceiver coil for human head imaging [29]. The resonator consisted in a 30-turn spiral coil made of Litz wire which hemispherical shape nicely fits the human head (height: 225 mm, width: 180 mm, depth: 100 mm). It was tuned and matched at 276 kHz, with a quality factor Q = 30 corresponding to a 10-kHz bandwidth. The most recent version of Boston's group low-field Halbach scanner reported by McDaniel and colleagues uses a similar single-channel helmet-shaped transceiver solenoid with a resistively-broadened 3-dB bandwidth of 78 kHz [58]. In their most recent compact cap design MRI system, McDaniel et al. [24] also presented a single-channel spiral-volume inspired transceiver coil. Wound inside a 3D printed former, a 4-turn coil made of Litz wire was resonated at 2.67 MHz, with its 3-dB bandwidth resistively broadened to reach 157 kHz (Q = 17). In their prototype MEG compatible resistive system, Galante et al. [56] describe the use of two separates coils geometrically decoupled for transmit and receive operations at 373 kHz. Their detection coil is a 27-turn saddle coil made of Litz wire wound on an 8-cm diameter cylinder, with a Q of 105 and corresponding bandwidth of 3.5 kHz. The receiver preamplifier located inside the MSR is running on battery to mitigate potential noise from the supply line. For their neonatal low-field system, Lother et al. [28] built a 10-cm inner diameter, 965 kHz transceiver solenoid coil. The design consisted of two parallel Litz wire assemblies made of 45 strands each. A quality factor Q = 95 with corresponding bandwidth BW = 10 kHz was measured for the coil placed inside the magnet. For their head imager, O'Reilly and colleagues used an 18-turn transceiver solenoid (diameter: 200 mm, length: 29 mm) made of copper wire at 2.15 MHz with 154-kHz bandwidth [23]. The authors also mention the use of an RF shield placed between their RF and gradient coils for an improved SNR of 9%. In their miniature, hand-held MRI system, Greer and colleagues have used a planar spiral design transceiver coil printed on two PCB layers [25]. The

A

Phantom

< 1 mT

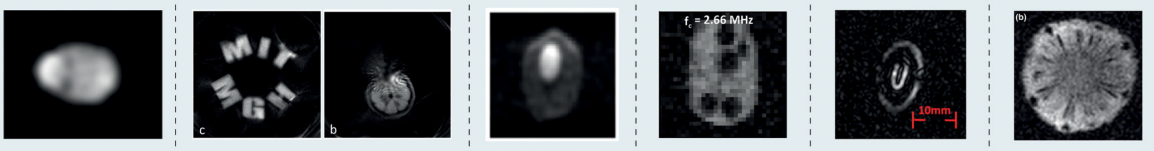


Technology	Prepolarization / SQUID	Prepolarization / SQUID	Prepolarization / SQUID	Prepolarization / OPAM	Prepolarization / Inductive	Prepolarization / Inductive
Magnet	Electromagnet (MSR)	Electromagnet (CMSB)	Electromagnet (no MSR)	Electromagnet (MSR)	Electromagnet (no MSR)	Electromagnet (no MSR)
B_p	210 mT	1.1 T	650 mT	65 mT	18.8 mT	18.8 mT
B_0	94 μT	93.7 μT	129 μT	117 μT	47 μT	47 μT
Acq. Matrix	32 x 16 (2D, 1 projection)	512 x 24 spokes (2D, 1 projection)	N x 31 (2D, 1 projection)	Not available (2D, 1 projection)	32 x 32 (2D, 1 projection)	32 x 32 (2D, 1 projection)
Rec. Matrix	64 x 64	Not available	Not available	Not available	Not available	Not available
Rec. Resol.	1.0 x 1.0 mm² (estimated)	0.662 x 0.662 x 8.000 mm³	Not available	7.4 x 8.9 mm²	3.125 x 3.125 x 170 mm³	3.125 x 3.125
Acq. Time	> 8 minutes (NA=8)	~ 3 minutes (estimated)	~ 17 minutes (NA=5, estimated)	> 10 min (NA=10, estimated)	8 min 32 s (NA=4)	Not available
SNR	~ 4 (after zero-filling, estimated)	Not available	Not available	Not available	Not available	Not available
Sample	Structured phantom	Structured phantom	Structured phantom	Water bottle	Cylinder with multiple-layer membrane	Structured phantom
Reference	Demachi, 2019 [49]	Kawagoe, 2016 [48]	Liu, 2015 [50]	Hilschenz, 2017 [55]	Ujihara, 2018 [33]	Benli, 2015 [32]

B

Phantom

[1:199] mT



Technology	No prepolarization / Inductive	No prepolarization / Inductive	No prepolarization / Inductive	No prepolarization / Inductive	No prepolarization / Inductive	No prepolarization / Inductive
Magnet	Electromagnet (MSR)	Permanent (no MSR)	Permanent (no MSR)	Permanent (no MSR)	Permanent (no MSR)	Electromagnet (w & w/o MSR)
B_p	none	none	none	none	none	none
B_0	8.9 mT	77.3 mT	50.4 mT	63.6 mT	186 mT	23 mT
Acq. Matrix	32 x 32 x 32 (3D)	256 x 91 rotations (2D, 1 projection) 256 x 181 rotations	64 x 64 x 64 (3D)	31 x 31 x 6 (2D multi-slice)	41 x 41 (2D, 1 slice)	64 x 128 (2D, 1 slice)
Rec. Matrix	Not available	256 x 256	64 x 64 x 64	31 x 31 x 6	256 x 256	128 x 128
Rec. Resol.	3 x 3 x 3 mm³	0.625 x 0.625 mm² (c: 1.7-cm, b: 1.0-cm thickness)	3.5 x 3.5 x 3.5 mm³	2.0 x 2.0 x 6.0 mm³	0.33 x 0.33 x 0.70 mm³	0.8 x 0.8 x 5.0 mm³
Acq. Time	2.3 hours (NA=16)	66 minutes 93 minutes	~ 34 minutes (NA=1)	12 min 40 s (NA=24)	54 minutes (NA=8)	29 minutes (NA=32)
SNR	149	Not available Not available	35 (after Gauss. filtering)	Not available	~10 (after zero-filling)	23 (after zero-filling)
Sample	Rabit brain, *ex vivo*	Structured phantom Lemon slice	Avocado in watermelon	Structured phantom	Structured phantom	Baby pineapple
Reference	Galante, 2015 [56]	Cooley, 2015 [21]	O'Reilly, 2019 [23]	McDaniel, 2019 [24]	Greer, 2019 [25]	Lother, 2016 [28]

FIGURE 3 | Summary of the current landscape of low-field MR imaging in phantoms. **(A)** Shows the major achievements obtained at ultra-low field (micro tesla range) in combination with SQUID detectors, and **(B)** the images obtained in the mT range (from 1 to 199 mT). MSR, magnetically shielded room; CMSB, compact magnetically shielded box; B_p, polarization field; NA, number of signal averaging. Images reused with permission from Cooley et al. [21], O'Reilly et al. [23], McDaniel et al. [24], Greer et al. [25], Lother et al. [28], Benli et al. [32], Ujihara et al. [33], Kawagoe et al. [48], Demachi et al. [49], Liu et al. [50], Hilschenz et al. [55]. Images reused from Galante et al. [56], held under Creative Commons License CC-BY 4.0.

coil is made of 3 turns per layer with inner diameter 9 mm and outer diameter 13 mm, added with a slotted shield placed over the top to limit external interferences. The resulting altered quality factor in presence of the shield is Q = 13.4 with corresponding bandwidth ∼600 kHz at 8 MHz Larmor frequency.

IMAGING PERFORMANCE

Phantom Studies

Imaging results in phantoms illustrate the potential of recent technological developments, from hardware (magnetics, RF, amplification) to software (sequence design, signal processing), even if not mature enough to be envisioned *in vivo*. **Figure 3** compiles all of the reported and most recent low- and ultra-low field related work.

MRI <1 mT

When considering ULF MRI with ultra-sensitive magnetometers, we can find more material from SQUID-based imaging as the latter technology is more mature. The work of Demachi et al. [49] shows pre-polarized 2D imaging (projection) at 4 kHz in a phantom composed of water and oil. From the parameters given it can be estimated that the minimum acquisition time was >8 min for a 2D 32 × 16 matrix interpolated to 64 × 64 in a phantom presenting four cylindrical columns (8-mm diameter, 19-mm depth), with a minimum polarization time of 0.125 s at B_p = 210 mT. The images displayed have extremely poor SNR. With a similar interest in food inspection application, Kawagoe et al. [48] show 2D imaging (projection) at ∼94 μT (∼4 kHz) in a 35-mm diameter, 8-mm thick phantom. A non-Cartesian radial acquisition scheme is used with 24 spokes, 512-ms acquisition time, and 5-s pre-polarization steps at B_p = 1.1 T (the highest reported here). The total acquisition time is not reported but could be estimated to ∼3 min for an interpolated, reconstructed voxel size of 662 × 662 × 8,000 μm^3. Liu et al. [50] show the feasibility of imaging without an MSR in a 30 × 38 mm^2 (projection, no depth given) structured phantom at 129 μT (5.5 kHz). From the given image parameters, it can be estimated that the imaging time was ∼17 min for their shortest phase encoding and acquisition times with a 5-s polarization at B_p =

650 mT. Cartesian acquisition was used in a spin-echo sequence with 31 phase-encode steps and number of average NA = 5 (no pixel size is given). The work of Hilschenz and colleagues shows 2D imaging (projection) with an OPAM operating at 117 μT (5 kHz) [55]. From a standard gradient echo approach following a 3-s pre-polarization step (B_p = 65 mT) and a total of NA = 10 averages, the team imaged the cross-section of a 34 mm-diameter bottle of di-ionized water. The given in-plane resolution of the 2D projection is 7.4 × 8.9 mm^2. SNR is not given, yet it looks quite poor. The total imaging time is not provided but can reasonably be expected to be above 10 min for a single projection from the parameters available. Benli [32] and Ujihara [33] have used inductive detection at 47 μT (∼2 kHz) and a 2D spin-echo sequence (projections) in their commercial benchtop system. Ujihara and colleagues performed imaging in 8 min 32 s with echo times TE ranging from 200 to 800 ms, TR = 4000 ms, BW = 32 Hz, NA = 4, matrix size 32 × 32, FoV = 100 × 100 mm^2, for a reconstructed pixel size of 3.125 × 3.125 mm^2. The pre-polarization time was 2 s at 18.8 mT. Information on SNR is not being communicated. Benli and colleagues have shown 2D T_2-weighted images with similar matrix size, FoV, and BW, but with TE ranging from 140 to 400 ms, TR = 5 s, and 4.5 s pre-polarization steps. The number of averages and total acquisition times are not being communicated. The SNR appears rather poor but is not being communicated.

MRI From 1mT to 199 mT

Sitting at constant B_0 = 8.9 mT static field (373 kHz), Galante and colleagues have shown MRI compatible with MEG settings in an MSR, but without pre-polarization [56]. Typical 3D Cartesian, spin-echo based acquisitions were performed with 32 × 32 × 32 matrix size, TE/TR = 19/500 ms, and a readout bandwidth in the kHz range. Scans in a doped water phantom and an *ex vivo* rabbit brain were acquired with 3 × 3 × 3 mm^3 voxel size, NA = 1 in 8.5 min, and NA = 16 in 2.3 h, respectively. SNR of 70 is reported for NA = 1 in the phantom. SNR of 149 is reported for NA = 16 *ex vivo*. A 2D spin-echo based projection was also acquired with 32 × 32 matrix size and down to 1 × 1 × 1 mm^3 voxel size in a ∼3-cm diameter, ∼10-mm high phantom filled with doped water. The latter shows rather poor SNR but no value is given. In their first attempt using a Halbach magnet geometry, Zimmerman-Cooley and colleagues were using the native field gradient from their magnet as a mean to perform spatial encoding [21]. The main advantage of such an approach is the simplicity of the design that does not require additional gradient or power electronics, thus facilitating flexibility and portability. The latter magnetic field gradients however are maximized in the periphery of the field of view with little or no gradient in the center, thus providing no spatial encoding (see **Figure 3B**). For their second iteration, the Boston team designed their magnet with an embedded linear gradient across the whole FoV, allowing spatial encoding from physically rotating the magnet around the object (multiple projections) [22], or by the addition of a gradient coil to modulate the latter in plane and sweep across *k*-space without rotation [58]. O'Reilly et al. [23] also report on imaging at 50.4 mT (2.15 MHz) using a custom Halbach magnet geometry. They acquired 3D images on a phantom composed of an avocado placed in a watermelon using a spin-echo sequence. Image parameters consisted in a 64 × 64 × 64 matrix with 220 × 220 × 220 mm^3 FoV, resulting in a voxel size of ∼3.5 × 3.5 × 3.5 mm^3. The acquisition bandwidth was 20 kHz, with a TE/TR = 30/500 ms, and NA = 1. The resulting acquisition time was TA = ∼34 min. Eventually, *k*−space data was filtered with a Gaussian filter before Fourier transform. The SNR in the images was reported to be ∼35. In their compact MR cap geometry, McDaniel et al. [24] demonstrate 2D multi-slice imaging at 63.6 mT (2.67 MHz) in a 4-cm high, 6.3-cm wide structured phantom which size matches their targeted ROI. Six slices were acquired using a slice-interleaved RARE sequence. Acquisition matrix was 31 × 31 for a ∼2 × 2 mm^2, 6-mm deep pixel resolution. The acquisition bandwidth was ∼54 kHz, TR = 1,000 ms, echo spacing of 3 ms, and NA = 24 for a total acquisition time of 12 min 40 s. Greer et al. [25] in their portable device made of 5 permanent magnets also ventured into imaging at 186 mT (8 MHz). They could show 2D imaging (1 selected slice) from a spin-echo based sequence as described elsewhere [59]. Typical parameters were 41 × 41 2D acquisition matrices with centric ordering of *k*−space, interpolated to 256 × 256 for a default slice thickness of ∼0.7 mm, and a pixel resolution of 0.33 × 0.33 mm^2. Reported SNR was ∼ 10 with acquisition time TA = 54 min and NA = 8. Signal intensity changes from the highly inhomogeneous B_0 are corrected by dividing each image with a reference scan. The overall images do represent well the imaged objects, yet with severe geometric distortions. Relying on an electromagnet design, Lother et al. [28] have shown imaging at 23 mT (965 kHz) in their custom compact MRI system dedicated to neonates. 2D Spin-echo over a 5-mm thick slice is shown in a baby pineapple. The imaging parameters of the sequence used were TE/TR = 40/400 ms, FoV = 100 × 200 mm^2, matrix size 64 × 128, BW = 100 Hz, and averaging NA = 32 for a total acquisition time TA = 29 min. Interpolation from zero-filling in *k*−space allows to shrink the displayed pixels from ∼1.6 × 1.6 mm^2 down to 0.8 × 0.8 mm^2. An SNR of 23 is reported on the displayed images.

In vivo Studies

Imaging *in vivo* applies to technology having reached a maturity level where most potential issues, hardware or software, are addressed to operate seamlessly in a living organism. The latter hence provides a better benchmark to assess performance such as SNR, contrast, and acquisition times in realistic objects and FoVs for future applications, being it clinical or pre-clinical. **Figure 4** compiles all of the reported most recent low- and ultra-low field work *in vivo*.

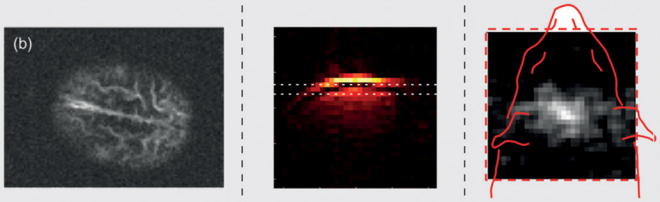
MRI <1 mT

In the range under 1 mT (< 43 kHz), the most advanced work presenting *in vivo* imaging in humans was obtained using SQUIDs inside of an MSR [38]. In the latter, pre-polarized imaging in the human head was performed at 200 μT (8 kHz), with voxel size 2 × 2.4 × 15 mm^3 using a 3D spin-echo sequence that includes 4 s pre-polarization steps at B_p = 100 mT, followed by a ∼100-ms ramp before spatial encoding starts. The total imaging time was 67 min for the acquisition a 5-slice volume. The SNR on the 2nd slice is reported to be ∼10, yet it drops

A

In vivo

< 1 mT

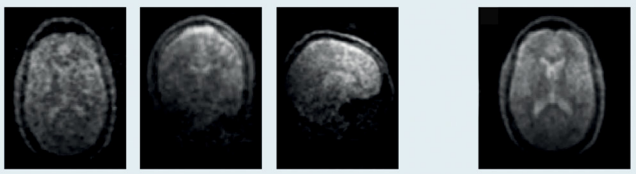


Technology	Prepolarization / SQUID	Prepolarization / SQUID	Prepolarization / SQUID
Magnet	Electromagnet (MSR)	Electromagnet (MSR)	Electromagnet (MSR)
Bp	100 mT	17 mT	10.6 mT
B0	200 μT	~ 39 μT	33 μT
Acq. Matrix	128 x 91 x 5 (3D)	N x 35 x 35 (3D)	32 x 32 x 8 (3D)
Rec. Matrix	Not available	N x 35 x 35	32 x 32 x 8
Rec. Resol.	2 x 2.4 x 15 mm^3	4.1 x 3.9 x 3.9 mm^3	1.3 x 1.3 x 2.6 mm^3
Acq. Time	67 minutes	~ 40 minutes	68 minutes (NA=8)
SNR	~ 10 on displayed slice	Not available	Not available
Sample	Human brain	Human brain	Rat head
Reference	Espy, 2015 [38]	Hömmen, 2019 [45]	Oyama, 2015 [47]

B

In vivo

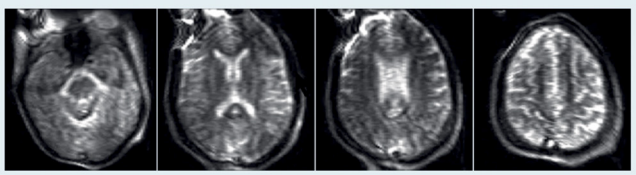
[1:199] mT



Technology	No prepolarization / Inductive			
Magnet	Electromagnet (no MSR)			
B_p	none			
B_0	6.5 mT			
Acq. Matrix	64 x 75 x 15 (3D with 50% undersampling in phase encoding directions)			
Rec. Matrix	64 x 75 x 15			
Rec. Resol.	2.5 x 3.5 x 8.5 mm^3			
Acq. Time	6 minutes (NA=30)	6 minutes (NA=30)	6 minutes (NA=30)	32 minutes (NA=160)
SNR	15	21	16	~ 40
Sample	Humain brain			
Reference	Sarracanie, 2015 [29]			

In vivo

[1:199] mT



Technology	No prepolarization / Inductive
Magnet	Permanent (no MSR)
B_p	none
B_0	79.3 mT
Acq. Matrix	254 x 49 x 23 (3D)
Rec. Matrix	Not available
Rec. Resol.	1.3 x 4.3 x 7.9 mm^3
Acq. Time	20 minutes (NA=8)
SNR	Not available
Sample	Human brain
Reference	McDaniel, 2019 [58]

FIGURE 4 | Summary of the current landscape of low-field MR imaging *in vivo*. **(A)** Shows the major achievements obtained at ultra-low field (micro tesla range) in combination with SQUID detectors, and **(B)** the images obtained in the mT range (from 1 to 199 mT). MSR, magnetically shielded room; CMSB, compact magnetically shielded box; B_p, polarization field; NA, number of signal averaging. Images reused, with permission, from Espy et al. [38], Hömmen et al. [45], Oyama et al. [47], McDaniel et al. [58]. Images reused from Sarracanie et al. [29], held under Creative Common License CC-BY 4.0.

drastically across the acquired volume. Similar in performance, Hömmen et al. [45] have performed pre-polarized imaging *in vivo* at lower field strength ∼39 μT (1,645 Hz) in the human brain with 4.1 × 3.9 × 3.9 mm^3 voxel size, from 35 $k-$steps in both 2nd and 3rd phase encoding directions. The sequence used was a 3D gradient-echo sequence with 500-ms pre-polarization steps at B_p = 17 mT, for a total acquisition time TA = 40 min. Only single slices are shown in 3 different orientations and SNRs are not reported. Finally, the work of Oyama shows imaging in a rat's head [47] at 33 μT (1.4 kHz) using 1 s pre-polarization steps at 10.6 mT peak field, and a 3D gradient echo acquisition with 32 × 32 × 8 matrix size. The total acquisition time was TA = 68 min for an 8-average scan (NA = 8) and voxel size 1.3 × 1.3 × 2.6 mm^3. 2D images exhibit rather low SNR although values are not provided (cf **Figure 4A**).

MRI From 1 mT to 199 mT

Sitting at the very bottom of the range (6.5 mT or 276 kHz), Sarracanie and colleagues could show images in the living human brain [29] by making use of high efficiency balanced steady state free precession sequences. They report 3D 64 × 75 × 15 acquisition matrices with voxel resolution down to 2.5 × 3.5 × 8.5 mm^3. Other parameters were TE/TR = 11/22.5 ms, number of averages NA = 30/160 and corresponding acquisition times TA = 6/32 min, using a 50% sampling rate with a variable density Gaussian pattern. The maximum SNR was reported to reach 15 (NA = 30) and ∼40 (NA = 160) using the latter parameters. In the continuity of Zimmerman-Cooley's work in phantoms, McDaniel and colleagues reported on imaging in the living human head at 79 mT (∼3.4 MHz) [58, 60]. Spin-echo sequences were employed using phase encoding for the 2nd and 3rd encoding spatial directions, along with frequency swept RF pulses so to provide coverage of nuclear spin frequencies over large magnetic field inhomogeneities [57]. They report 254 × 49 × 23 acquisition matrices providing 1.3 × 4.3 × 7.9 mm^3 voxel size, with NA = 8, TE/TR = 148/3000 ms, and read-out bandwidth = 100 kHz, for a total acquisition time TA = 20 min. Efforts were made to passively shim the magnet from specific trays embedded in the original design, and a generalized reconstruction was performed and confronted to fast Fourier transform. The generalized reconstruction performs better than traditional Fourier, and although very encouraging, the reconstructed images exhibits some heavy distortions, in particular on the edges of the object where the magnetic field is most heterogeneous (see **Figure 4B**).

SIDETRACKS

The authors have tried to keep the extent of the present paper within a certain category of academic work and a specific time period, so to depict a refreshed view on the topic while keeping coherence for further discussion. Nonetheless, it is worth mentioning potent initiatives on the edge of our scope as some other contenders propose truly original approaches to MRI that also leverage low magnetic field strengths. At field strength ≥ 0.2 T, one start to enter the realm of low-field commercial systems less in line with the scope of the proposed overview. With a particular focus on mobility and accessibility, Nakagomi and colleagues have proposed a prototype, car-mounted MRI system operating at 0.2 T (8.5 MHz) [61]. Relying on a bi-planar permanent magnet architecture with a set of bi-planar gradient coils for spatial encoding, the team demonstrates 2D multi-slice imaging in the elbow within ∼1 min 30 s. To the best of our knowledge, this is the first time that a full MRI system for human imaging is being sited in a standard, commercial vehicle. This initiative, even at the stage of prototyping, advocates strongly for the use of lower field technology to promote flexibility in future mobile systems. Similarly, low static magnetic field (0.35 T or 14.9 MHz) was successfully employed in the first CE marked (and FDA cleared) MRI-guided radiation therapy cancer treatment system by ViewRay (MRIdian Linac, ViewRay Inc., Oakwood, USA) [62]. It further demonstrates that bringing down magnetic field is a relevant option to combine complex and rather incompatible modalities within the same device. Balanced steady state (b-SSFP) based MRI sequences are used that are particularly well-suited to low-field. From an increased $\frac{T_2}{T_1}$ ratio, the transverse magnetization at steady-state is increased that promotes SNR [63], and from lower sensitivity to magnetic field inhomogeneity, the images acquired are less prone to typical banding artifact encountered at higher field [29]. Another growing interest in MRI lies in the acquisition of quantified metrics to replace the long-lasting legacy of reading shades of gray. Field cycling relaxometry is a technique that started to be used in NMR spectroscopy, which consists in measuring the longitudinal relaxation rate T_1 in samples at multiple field strengths [64–67]. T_2 relaxation would probably need some deeper investigation in the low-field range, but is currently less of interest due to little dependence with field strength [68]. T_1 on the contrary is known to exhibit clear changes. Its evolution as a function of field is not linear and hence can be used as an extra degree of freedom to characterize tissue types, and potentially help in the identification of new biomarkers for a given pathology. At low field in particular, the dispersion in T_1 relaxation rates is much bigger than at clinical field strength (or higher) and moves away from that of pure water as more and more molecular motion contribute to relaxation processes [69]. Since it is commonly accepted that the water content of tissue is neither tissue nor disease dependent, the MR community probably should embrace the exploration of new paths for diagnosis purpose. Regarding its application in MRI however, relaxometry is not used in clinical routine due to the additional time needed to probe the NMR signal at multiple time points. Accounting for the multiple time points necessary, the multiple static magnetic field strength to be investigated, and the loss of SNR intrinsic to lower proton magnetization, it might be hard to imagine field cycling approaches being implemented with future clinical low field MRI systems. Despite these challenges, a group of researchers from the University of Aberdeen has however managed to build the first whole body Fast Field-Cycling MRI scanner and used it for clinical molecular imaging studies [70]. With all legal authorizations to conduct studies in human volunteers and patients, the team has managed to provide T_1 dispersion maps *in vivo* in the living human breast, brain and knee that already points toward contrasts and diagnostic

capability unique to low field. The total duration of FFC-MRI scans varied between 35 and 50 min from positioning of the patient to withdrawal. With a constant detection field of 0.2 T (8.5 MHz) higher than reported in this paper, the range of investigated T_1s spans nonetheless from 50 μT to 0.2 T, hence very low field strengths.

COMMERCIAL PERSPECTIVES

If the research at very low-field strength seems to be quite dynamic again, it is still hard to predict what will eventually result in commercial solutions for clinical practice in the long term. In order to add value and potentially find new markets, it is reasonable to assume that such devices will not be designed to compete with high-field settings, but instead address current needs where they can be complementary and accessible. Amongst the main avenues considered, systems to be sited in emergency units, in the operating room or even in the field will likely be developed for applications in neurology or musculoskeletal disorders, where some well-defined needs already exist (e.g. stroke diagnosis). With a permanent, bi-planar magnet architecture operating at 64 mT (∼2.7 MHz), the first and only contender to a commercial, accessible and mobile MRI comes from Hyperfine (Guilford, CT, USA), an American start-up company that just unveiled their MRI system in late 2019. The whole MRI system and embedded electronics is mounted on a cart with motorized wheels and can be plugged into a standard wall outlet. It has not yet been cleared with CE approval, but was recently granted 510(k) FDA clearance and is currently being tested in several clinical settings on the east-coast of the United States. With no specific need for siting and low power requirement, it can easily be placed in the emergency department, the neuro-intensive care or pediatric units. It was first intended for neuro-imaging and muskulo-skeletal applications. At a different scale, it is worth mentioning that the main manufacturers also seem to be reconsidering their position toward low-field MRI. In their first attempt, low-field was only considered for open access permanent magnet based devices, which reduced cost did not allow to increase the overall value, mainly due to important siting requirements and disadvantaged by overall inferior performance. To the best of our knowledge, no commercial device is foreseen yet, but a recent publication has communicated about one of the main manufacturers (Siemens Healthineers, Erlangen, Germany) ramping a clinical 1.5 T (∼64 MHz) scanner down to 0.55 T (23.4 MHz) for a variety of applications [15]. Benefiting from lower sensitivity to magnetic susceptibility effects and enhanced contrast in the mid-field regime, it has resulted in a series of preliminary results recently communicated by the same group [71–74].

DISCUSSION

In this paper, a rather extensive review of the current landscape of low to ultra-low field MRI has been listed and described. We decided to focus on the last 5 years of development to give a fresh view on the topic, though we encourage the reader to learn more about older inspiring pioneer work in the field. Indeed, very interesting research has been published especially in the late 80s—early 90s [1, 4, 75–77], with a brief recurring interest in the early 2000s, in particular for interventional applications [13, 14]. The MR community is clearly entering one of those cycles where the interest in low- and ultra-low field MRI is raising again, and the past few years have shown a remarkable increase of communications regarding new hardware and imaging techniques. In this context, we try to present an exhaustive list of approaches reflecting the current research effort, all of which is capable of generating images at field strengths as low as a few μT. From this "snapshot" of the current state of the art, it is possible to take some perspective and discuss what seems to be most relevant, at least in the short and medium term, or what kind of technological locks would need to be addressed in order for some approaches to work. Indeed, not all of the work presented has a potential to translate into *in vivo* and further into clinical applications, the main reason being that such settings require simple-to-use, rugged and rather fast technologies. The authors anticipate that pre-polarizing approaches that are quite well-represented in low field research might have difficulties to meet clinical needs both on the infrastructure and performance levels. It is indeed quite striking to observe that when using pre-polarization, often more than 95% of the examination time is being used field cycling and not acquiring data. We can notice that while sensitivity is already quite reduced from working at weaker field strength, the low duty cycle on the acquisition side is never quite compensated by a higher (even very high [48]) pre-polarizing field or higher sensitivity sensors. One might consider using that time to acquire and average data as a more efficient way to increase SNR, with the added benefit of simpler setups if field cycling is not required. In addition, the longer readout and encoding times generally encountered added with incompressible delays to cycle down the pre-polarization fields also penalize signals with short lifetimes, and one may end up only being able to probe species with long relaxation time such as free water compartments as seen with Espy and colleagues [38] (**Figure 4A**). Pre-polarization techniques are often employed for magneto-encephalography in conjunction with high sensitivity magnetometers but new insights for alternatives are proposed, such as in the work from Galante [31, 56]. Imaging at thermal equilibrium without pre-polarization seems to be the most promising approach considering the image quality achievable today. Permanent and resistive magnets are two good candidates that have both pros and cons. Permanent magnet architectures have the big advantage of requiring no power, key to the prospect of future small footprint, accessible technology. The achievable field homogeneity however is easily one to two orders of magnitude worse than what can be found in resistive magnets. As a result, strong artifacts that impede image quality can be seen in all of the studies presented (**Figures 3**, **4**), that are very hard to correct. Innovative reconstruction methods will be required to overcome this difficulty, probably moving away from typical Fourier frameworks, and leveraging for example deep learning approaches. If this technical challenge has not yet been unlocked, addressing this reconstruction issue has the potential to truly disrupt the field. Permanent magnets also suffer from poor field stability with respect to temperature, which certainly interrogates on their deployment in extreme

environments without controlled temperature and humidity. On the other hand, resistive magnet architectures provide better field homogeneity and the ability to change the desired field of operation or even switch it off, but at the cost of higher power needs as well as cooling. Then, depending on the targeted magnet size and field strength, simple cooling solutions are foreseeable (forced air vs. water) before requiring any complex or demanding cooling resources. Eventually, both permanent and resistive magnet technologies are worth exploring when it comes to low and ultra-low field imaging, keeping in mind the requirements and constraints for the targeted applications and corresponding working environments. A general comment in view of the multiple attempts to build smaller magnets could be to assess imaging capability first in a controlled setting before construction is envisioned. That could be an efficient way to save time on such developments, rather than empirically iterating on complex magnet construction and later realize whether or not imaging is feasible. Regarding field strength, it is quite interesting to note that all the attempts to low-field imaging reported here cover quite a broad range, and that a higher field strength does not bring superior image quality (**Figures 3**, **4**). Overall, the range 5–100 mT appears to be the most promising in delivering fast imaging *in vivo*, at both extremities of the spectrum [24, 29]. From the latter results, we believe that the democratization of low field MRI will not come from reaching higher fields, but instead being able to stay low. The key is to embrace the advantages offered by this unique regime while gathering a set of technological solutions that maximize SNR/unit time (i.e., from sensors, acquisition schemes) and navigate through magnetic field inhomogeneity. Further, we believe that approaches such as model-based MRI and deep learning will appear of upmost relevance to navigate through poor SNR, while promoting access to simpler hardware and hence lower cost and reduced physical footprint. Another major and often overlooked feature of low-field MRI deserves further discussion: contrast. It is known from the early days of MRI that low-field strengths provide a higher dispersion in T_1 relaxation times [78] (see **Figure 5**) but the latter never really was extensively explored. Many scientists may be working with "standard" high-field commercial equipment and completely be missing which field strength provides the best contrast (i.e., T_1, T_2 dispersion) with respect to their individual interests. An interesting change of paradigm could be envisioned where exploratory studies would be performed in dedicated MR systems, ramping the field in order to assess the dispersion in relaxation rates and potentially finding the sweet spot for maximum contrast. Eventually, future generations of low-field MRI systems with field cycling possibilities could be anticipated where contrast can be tuned to a specific application. The latter is not too far away from current fast-field cycling MRI initiatives described above in this manuscript, aimed at uncovering T_1 dispersion as an intrinsic metric of interest in MRI diagnosis [70]. Naturally, such considerations on contrast open new perspectives regarding diagnosis capability, yet it will also challenge radiologists to adapt their skills in the interpretation of images according to the field strength of operation. It may also be in the hands of other practitioners to complement radiologists and develop basic skills at reading images (as many already do), such that new class of point-of-care units truly succeed to decongest radiology departments. Alternatively, one may see these devices more as tools to answer simple clinical questions, rather than a complex imaging device that belong to the radiology department. Coming back to the title of this paper and its question *how low can we go?* our experience suggests that it is likely to be an ill-posed problem. The question should be application-driven and address specific goals as well as technical constraints. Let us go back in time and make an analogy with aerospace research. The question in the 1960s was: "what technological leap is needed to go on the moon?" and not "what technology should we build in order to explore the entire galaxy?" In medical imaging, scientists and clinicians have adapted ultrasound or X-ray-based modalities to address challenges in different disciplines of medicine. MRI on the other hand has been almost untouched for the last three decades and one-fits-all scanners continue to be the main stream. As a consequence, MRI has become a bottleneck in the hospital care chain for 50% of the total world population, and the pecuniary burden linked to scanner purchase, siting and maintenance makes it inaccessible for the remaining 50% (see **Figure 6**). Indeed, cost is undeniably an issue in the accessibility of MRI technology. If financial support can always be found for a one-time purchase, it is difficult to find resources for costly maintenance contracts or simply to maintain infrastructure over long time periods, as infrastructure can also be of critical importance. Examples include recurring power outages, either because of lack of financial means in developing countries, or in war zones. Electricity blackouts can severely damage medical devices that will never be repaired and pile-up in endless medical device graveyards. Leveraging low magnetic field in scanners addressing these difficulties would certainly open a market in such regions, very similar to what the start-up company Pristem SA did with their robust and low-cost X-ray system GlobalDiagnostiX (Pristem SA, Lausanne,

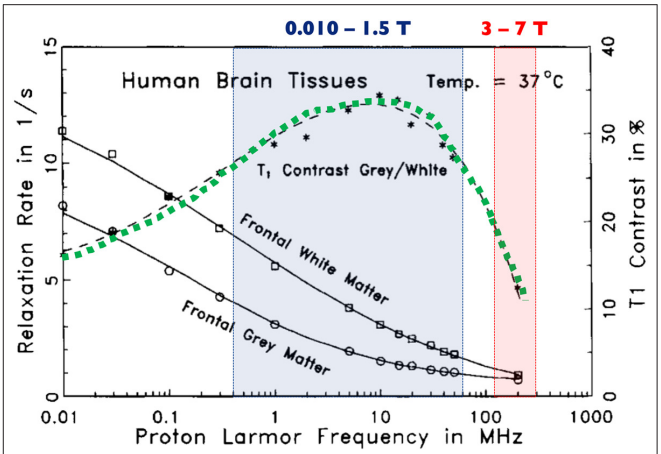


FIGURE 5 | Relative T_1 contrast in human brain samples as a function of Larmor frequency. The green dotted line shows that the relative contrast between gray and white matter is higher when operating between 10 mT and 1.5 T (shaded blue box) and decreases dramatically at very high to ultra-high field strengths (shaded red box). Modified, with permission, from Fischer et al. [78].

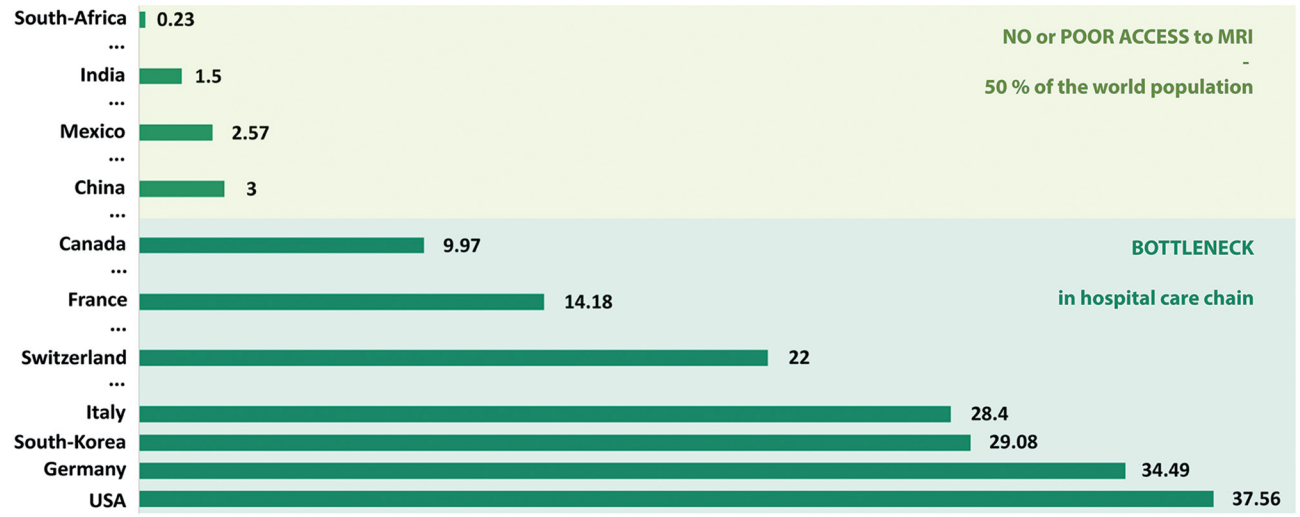


FIGURE 6 | Number of MRI units per million population. Only a sample of countries are presented here in order to simply illustrate the current MRI market worldwide. Data extracted mainly from Statistica.com, 2017 reports.

Switzerland). Accessibility is also bound to siting resources. Standard MRI units require three separate rooms: one for the operator, the RF and magnetically shielded examination room, and the technical room with all the associated power electronics. Highly populated regions, like most major cities in Asia, often have the means to buy MRI machines, but not to site them. Low field alternatives that do not need specific shielding or heavy power requirements certainly would address this challenge and again open an untouched market. For the wealthiest half of the world population, low-field MRI dedicated to a range or a specific application (neurology, MSK, etc.), has the potential to decongest radiology departments and also facilitate magnetic resonance in multi-modal settings from an intrinsic enhanced compatibility. Sited outside of the radiology department, such MRI units would become handy tools to the practitioner, accessible to all disciplines and population, and hence start to change paradigm from the historical one-fits-all. The same approach might ultimately follow at high and ultra-high field where the application envisioned will lead manufacturers to offer purpose-built systems that fit the needs of research and medical professionals. Unless one wants to address the underlying question: how high shall we go for high-field MRI?

CONCLUSION

Low field MRI has regained popularity over the past few years, reopening the debate on its relevance in the clinical setting. With half of the world population underserved regarding MR diagnosis, the medical community indeed strives for more ubiquitous and accessible systems. Lowering magnetic field strength opens new perspectives to increase MR value not only from reduced costs, but also from enhanced outcomes, shifting paradigm toward specific use cases and more adaptability. It promotes new magnet geometries that have already performed imaging *in vivo* in humans, and which low-footprint core technology is no longer associated with superconductive materials. Lower field MRI as such could diversify the current offer to a broader range of medical applications and geographical locations, with a wider range of contrasts to complement current diagnostic tools. As a conclusion, we believe that it is time to go as low as diagnostically relevant. It is the scientific and medical communities' responsibility to choose how low/high one should go depending on specific applications, and to work side-by-side with the industry to build the future of MRI diagnosis.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was supported by the Swiss National Science Foundation (grants #PP00P2_170575; PCEFP2_186861).

REFERENCES

1. Sepponen RE, Sipponen JT, Sivula A. Low-field (20mT) NMRI of the brain. *J Comput Assist Tomogr.* (1985) **9**:237–41. doi: 10.1097/00004728-198503000-00002
2. Sipponen JT, Sepponen RE, Tanttu JI, Sivula A. Intracranial hematomas studied by MRI at 0.17 and 0.02 T. *J Comput Assist Tomogr.* (1985) **9**:698–704. doi: 10.1097/00004728-198507010-00007
3. Hovi I, Korhola O, Valtonen M, Valtonen V, Taavitsainen M, Kivisaari A, et al. Detection of soft-tissue and skeletal infections with ultra low-field

(0.02 T) MRI. *Acta Radiol.* (1989) **30**:495–9. doi: 10.3109/02841858909175316
4. Constantinesco A, Arbogast S, Foucher G, Vinée P, Choquet P. Detection of glomus tumor of the finger by dedicated MRI at 0.1 T. *Magn Reson Imaging.* (1984) **12**:1131–4. doi: 10.1016/0730-725X(94)91246-S
5. Kaufman L. The Impact Of Radiology's Culture On The Cost Of Magnetic Resonance Imaging. *J Magn Reson Imaging.* (1996) **68**:67–71. doi: 10.1002/jmri.1880060113
6. Edelman RR. The History of MR imaging as seen through the pages of radiology. *Radiology.* (2014) **273**:S181–200. doi: 10.1148/radiol.14140706
7. Moser E, Laistler E, Schmitt F, Kontaxis G. Ultra-high field NMR and MRI—the role of magnet technology to increase sensitivity and specificity. *Front Phys.* (2017) **5**:33. doi: 10.1016/S0140-6736(11)60282-1
8. Zimmerman RA, Bilaniuk LT, Hackney DB, Goldberg HI, Grossman RI. Head injury: early results of comparing CT and high-field MR. *AJR.* (1986) **147**:1215–22. doi: 10.2214/ajr.147.6.1215
9. Choquet P, Breton E, Goetz C, Marin C, Constantinesco A. Dedicated low-field MRI in mice. *Phys Med Biol.* (2009) **54**:5287–99. doi: 10.1111/j.1740-8261.1995.tb00306.x
10. Rinck PA. MR imaging: quo vadis? *Rinckside.* (2019) **30**:3.
11. van Beek EJR, Kuhl C, Anzai Y, Desmond P, Ehman RL, Gong Q, et al. Value of MRI in medicine: more than just another test? *J Magn Reson Imaging.* (2019) **49**:e14–25. doi: 10.1148/radiol.2016150789
12. Iturri-Clavero F, Galbarriatu-Gutierrez L, Gonzalez-Uriarte A, Tamayo-Medel G, de Orte K, Martinez-Ruiz A, et al. "Low-field" intraoperative MRI: a new scenario, a new adaptation. *Clin Radiol.* (2016) **71**:1193–8. doi: 10.1016/j.crad.2016.07.003
13. Senft C, Ulrich CT, Seifert V, Gasser T. Intraoperative magnetic resonance imaging in the surgical treatment of cerebral metastases. *J Surg Oncol.* (2010) **10**:436–41. doi: 10.1007/s11060-009-9868-6
14. Senft C, Seifert V, Hermann E, Gasser T. Surgical treatment of cerebral abscess with the use of a mobile ultralow-field MRI. *Neurosurg Rev.* (2008) **32**:77–85. doi: 10.1016/0090-3019(93)90008-O
15. Campbell-Washburn AE, Ramasawmy R, Restivo MC, Bhattacharya I, Basar B, Herzka DA, et al. Opportunities in interventional and diagnostic imaging by using high-performance low-field-strength MRI. *Radiology.* (2019) **293**:384–93. doi: 10.1148/radiol.2019190452
16. Schukro C, Puchner SB. Safety and efficiency of low-field magnetic resonance imaging in patients with cardiac rhythm management devices. *Eur J Radiol.* (2019) **118**:96–100. doi: 10.1016/j.ejrad.2019.07.005
17. Klein HM. *Clinical Low Field Strength Magnetic Resonance Imaging*. Cham: Springer International Publishing AG Switzerland (2016). doi: 10.1007/978-3-319-16516-5
18. Hoult DI, Richards RE. The signal-to-noise ratio of the nuclear magnetic resonance experiment. *J Magn Reson.* (1976) **24**:71–85. doi: 10.1016/j.jmr.2011.09.018
19. Marques JP, Simonis FFJ, Webb AG. Low-field MRI: an MR physics perspective. *J Magn Reson Imaging.* (2018) **49**:1528–42. doi: 10.1002/mrm.26221
20. Darrasse L, Ginefri JC. Perspectives with cryogenic RF probes in biomedical MRI. *Biochimie.* (2003) **85**:915–37. doi: 10.1016/j.biochi.2003.09.016
21. Cooley CZ, Stockmann JP, Armstrong BD, Sarracanie M, Lev MH, Rosen MS, et al. Two-dimensional imaging in a lightweight portable MRI scanner without gradient coils. *Magn Reson Med.* (2015) **73**:872–83. doi: 10.1002/mrm.25147
22. Cooley CZ, Haskell MW, Cauley SF, Sappo C, LaPierre CD, Ha CG, et al. Design of sparse Halbach magnet arrays for portable MRI using a genetic algorithm. *IEEE Trans Magn.* (2018) **54**:1–12. doi: 10.1109/TMAG.2017.2751001
23. O'Reilly T, Teeuwisse WM, Webb AG. Three-dimensional MRI in a homogenous 27 cm diameter bore Halbach array magnet. *J Magn Reson.* (2019) **307**:106578. doi: 10.1016/j.jmr.2019.106578
24. McDaniel PC, Cooley CZ, Stockmann JP, Wald LL. The MR Cap: a single-sided MRI system designed for potential point-of-care limited field-of-view brain imaging. *Magn Reson Med.* (2019) **82**:1946–60. doi: 10.1006/jmre.2000.2263
25. Greer M, Chen C, Mandal S. An easily reproducible, hand-held, single-sided, MRI sensor. *J Magn Reson.* (2019) **308**:106591. doi: 10.1016/j.jmr.2019.106591
26. Eidmann G, Savelsberg R, Peter B, Blümich B. The NMR MOUSE, a mobile universal surface explorer. *J Magn Reson.* (1996) **122**:104–9. doi: 10.1006/jmra.1996.0185
27. Blümich B, Peter B, Guthausen A, Haken R, Schmitz U, Saito K, et al. The NMR-MOUSE: construction, excitation, and applications. *Magn Reson Imaging.* (1998) **16**:479–84. doi: 10.1016/S0730-725X(98)00069-1
28. Lother S, Schiff SJ, Neuberger T, Jakob PM, Fidler F. Design of a mobile, homogeneous, and efficient electromagnet with a large field of view for neonatal low-field MRI. *Magn Reson Mater Phys.* (2016) **29**:691–8. doi: 10.1007/s10334-016-0525-8
29. Sarracanie M, LaPierre CD, Salameh N, Waddington DEJ, Witzel T, Rosen MS. Low-cost high-performance MRI. *Sci Rep.* (2015) **5**:15177. doi: 10.1038/srep15177
30. Tsai LL, Mair RW, Rosen MS, Patz S, Walsworth RL. An open-access, very-low-field MRI system for posture-dependent 3He human lung imaging. *J Magn Reson.* (2008) **193**:274–85. doi: 10.1016/j.jmr.2008.05.016
31. Galante A, Catallo N, Sebastiani P, Sotgiu A, Sinibaldi R, De Luca C, et al. *Very Low Field MRI: a Fast System Compatible with Magnetoencephalography*. Turin: IEEE MeMeA Proc. (2015) 560–4. doi: 10.1109/MeMeA.2015.7145266
32. Benli KP, Dillmann B, Louelh R, Poirier-Quinot M, Darrasse L. Illustrating the quantum approach with an Earth magnetic field MRI. *Eur J Phys.* (2015) **36**:035032. doi: 10.1038/334019c0
33. Ujihara R, Fridjonsson EO, Bristow NW, Vogt SJ, Bucs SS, Vrouwenvelder JS, et al. Earth's field MRI for the non-invasive detection of fouling in spiral-wound membrane modules in pressure vessels during operation. *Desalin Water Treat.* (2018) **135**:16–24. doi: 10.5004/dwt.2018.23156
34. Zotev VS, Matlashov AN, Volegov PL, Savukov IM, Espy MA, Mosher JC, et al. Microtesla MRI of the human brain combined with MEG. *J Magn Reson.* (2008) **194**:115–20. doi: 10.1016/j.jmr.2008.06.007
35. Zotev VS, Volegov PL, Matlashov AN, Espy MA, Mosher JC, Kraus RH Jr. Parallel MRI at microtesla fields. *J Magn Reson.* (2008) **192**:197–208. doi: 10.1016/j.jmr.2008.02.015
36. Volegov P, Flynn M, Kraus R, Magnelind P, Matlashov A, Nath P, et al. Magnetic resonance relaxometry at low and ultra low fields. *IFMBE Proc.* (2010) **28**:82–7. doi: 10.1007/978-3-642-12197-5_15
37. Vesanen PT, Zevenhoven KCJ, Nieminen JO, Dabek J, Parkkonen LT, Ilmoniemi RJ. Temperature dependence of relaxation times and temperature mapping in ultra-low-field MRI. *J Magn Reson.* (2013) **235**:50–7. doi: 10.1016/j.jmr.2013.07.009
38. Espy MA, Magnelind PE, Matlashov AN, Newman SG, Sandin HJ, Schultz LJ, et al. Progress toward a deployable SQUID-based ultra-low field MRI system for anatomical imaging. *IEEE Trans Appl Supercond.* (2015) **25**:1–5. doi: 10.1109/TASC.2014.2365473
39. McDermott R, Lee S, Haken te B, Trabesinger AH, Pines A, Clarke J. Microtesla MRI with a superconducting quantum interference device. *Proc Natl Acad Sci USA.* (2004) **101**:7857–61. doi: 10.1073/pnas.0402382101
40. Inglis B, Buckenmaier K, SanGiorgio P, Pedersen AF, Nichols MA, Clarke J. MRI of the human brain at 130 microtesla. *Proc Natl Acad Sci USA.* (2013) **110**:19194–201. doi: 10.1073/pnas.1319334110
41. Ganssle PJ, Shin HD, Seltzer SJ, Bajaj VS, Ledbetter MP, Budker D, et al. Ultra-Low-Field NMR relaxation and diffusion measurements using an optical magnetometer. *Angew Chem.* (2014) **126**:9924–8. doi: 10.1063/1.3623024
42. Lin FH, Vesanen PT, Hsu YC, Nieminen JO, Zevenhoven KCJ, Dabek J, et al. Suppressing multi-channel ultra-low-field MRI measurement noise using data consistency and image sparsity. *PLoS ONE.* (2013) **8**:e61652. doi: 10.1371/journal.pone.0061652.g005
43. Clarke J, Hatridge M, Mößle M. SQUID-detected magnetic resonance imaging in microtesla fields. *Annu Rev Biomed Eng.* (2007) **9**:389–413. doi: 10.1146/annurev.bioeng.9.060906.152010
44. Dong H, Zhang Y, Krause HJ, Xie X, Offenhäusser A. Low field MRI detection with tuned HTS SQUID magnetometer. *IEEE Trans Appl Supercond.* (2011) **21**:509–13. doi: 10.1109/TASC.2010.2091713
45. Hömmen P, Storm JH, Höfner N, Körber R. Demonstration of full tensor current density imaging using ultra-low field MRI. *Magn Reson Imaging.* (2019) **60**:137–44. doi: 10.1016/j.mri.2019.03.010
46. Drung D, Assmann C, Beyer J, Kirste A, Peters M, Ruede F, et al. Highly sensitive and easy-to-use SQUID sensors. *IEEE Trans Appl Supercond.* (2007) **17**:699–704. doi: 10.1109/TASC.2007.897403

47. Oyama D, Higuchi M, Kawai J, Tsuyuguchi N, Miyamoto M, Adachi Y, et al. *Measurement of Magnetic Resonance Signal from a Rat Head in Ultra-low Magnetic Field*. Nagoya: ISEC Proc. (2015) 1–3. doi: 10.1109/ISEC.2015.7383470
48. Kawagoe S, Toyota H, Hatta J, Ariyoshi S, Tanaka S. Ultra-low field MRI food inspection system prototype. *Phys C.* (2016) **530**:104–8. doi: 10.1016/j.physc.2016.02.015
49. Demachi K, Hayashi K, Adachi S, Tanabe K, Tanaka S. T1-weighted image by ultra-low field SQUID-MRI. *IEEE Trans Appl Supercond.* (2019) **29**:1600905. doi: 10.1109/TASC.2019.2902772
50. Liu C, Chang B, Qiu L, Dong H, Qiu Y, Zhang Y, et al. Effect of magnetic field fluctuation on ultra-low field MRI measurements in the unshielded laboratory environment. *J Magn Reson.* (2015) **257**:8–14. doi: 10.1016/j.jmr.2015.04.014
51. Savukov I, Karaulanov T, Castro A, Volegov P, Matlashov A, Urbatis A, et al. Non-cryogenic anatomical imaging in ultra-low field regime: Hand MRI demonstration. *J Magn Reson.* (2011) **211**:101–8. doi: 10.1016/j.jmr.2011.05.011
52. Savukov I, Karaulanov T. Anatomical MRI with an atomic magnetometer. *J Magn Reson.* (2013) **231**:39–45. doi: 10.1016/j.jmr.2013.02.020
53. Savukov I, Karaulanov T, Wurden CJV, Schultz L. Non-cryogenic ultra-low field MRI of wrist–forearm area. *J Magn Reson.* (2013) **233**:103–6. doi: 10.1016/j.jmr.2013.05.012
54. Savukov I, Karaulanov T. Magnetic-resonance imaging of the human brain with an atomic magnetometer. *Appl Phys Lett.* (2013) **103**:043703. doi: 10.1109/TASC.2009.2018764
55. Hilschenz I, Ito Y, Natsukawa H, Oida T, Yamamoto T, Kobayashi T. Remote detected low-field MRI using an optically pumped atomic magnetometer combined with a liquid cooled pre-polarization coil. *J Magn Reson.* (2017) **274**:89–94. doi: 10.1016/j.jmr.2016.11.006
56. Galante A, Sinibaldi R, Conti A, De Luca C, Catallo N, Sebastiani P, et al. Fast room temperature very low field-magnetic resonance imaging system compatible with MagnetoEncephaloGraphy environment. *PLoS ONE.* (2015) **10**:e0142701. doi: 10.1371/journal.pone.0142701.g011
57. Stockmann JP, Cooley CZ, Guerin B, Rosen MS, Wald LL. Transmit array spatial encoding (TRASE) using broadband WURST pulses for RF spatial encoding in inhomogeneous B0 fields. *J Magn Reson.* (2016) **268**:36–48. doi: 10.1016/j.jmr.2016.04.005
58. McDaniel PC, Cooley CZ, Stockmann JP, Wald LL. A target-field shimming approach for improving the encoding performance of a lightweight Halbach magnet for portable brain MRI. In: *Proceedings of the International Society for Magnetic Resonance in Medicine.* Montreal (2019). p. 0215.
59. Perlo J, Casanova F, Blümich B. 3D imaging with single-sided sensor: an open tomograph. *J Magn Reson.* (2004) **166**:228–35. doi: 10.1016/j.jmr.2003.10.018
60. McDaniel PC, Cooley CZ, Stockmann JP, Wald LL. 3D imaging with a portable MRI scanner using an optimized rotating magnet and 1D gradient coil. In: *Proceedings of the International Society for Magnetic Resonance in Medicine.* Paris (2018). p. 0029.
61. Nakagomi M, Kajiwara M, Matsuzaki J, Tanabe K, Hoshiai S, Okamoto Y, et al. Development of a small car-mounted magnetic resonance imaging system for human elbows using a 0.2 T permanent magnet. *J Magn Reson.* (2019) **304**:1–6. doi: 10.1016/j.jmr.2019.04.017
62. Klüter S. Technical design and concept of a 0.35 T MR-Linac. *Clin Transl Radiat Oncol.* (2019) **18**:98–101. doi: 10.1016/j.ctro.2019.04.007
63. Scheffler K, Lehnhardt S. Principles and applications of balanced SSFP techniques. *Eur Radiol.* (2003) **13**:2409–18. doi: 10.1007/s00330-003-1957-x
64. Koenig SH, Hallenga K, Shporer M. Protein-water interaction studied by solvent 1H, 2H, and 17O magnetic relaxation. *Proc Natl Acad Sci USA.* (1975) **72**:2667–71. doi: 10.1073/pnas.72.7.2667
65. Grösch L, Noack F. NMR relaxation investigation of water mobility in aqueous bovine serum albumin solutions. *Biochim Biophys Acta BBA.* (1976) **453**:218–32. doi: 10.1016/0005-2795(76)90267-1
66. Kimmich R. Field cycling in NMR relaxation spectroscopy: applications in biological, chemicaland polymer physics. *Bull Magn Reson.* (1980) **1**:195–218.
67. Noack F. NMR field-cycling spectroscopy: principles and applications. *Prog Nucl Magn Reson Spectrosc.* (1986) **18**:171–276. doi: 10.1016/0079-6565(86)80004-8
68. Bottomley PA, Foster TH, Argersinger RE, Pfeifer LM. A review of normal tissue hydrogen NMR relaxation times and relaxation mechanisms from 1-100 MHz: dependence on tissue type, NMR frequency, temperature, species, excision, and age. *Med Phys.* (1998) **11**:425–48. doi: 10.1118/1.595535
69. Rinck PA, Fischer HW, Vander Elst L, Van Haverbeke Y, Muller RN. Field-cycling relaxometry: medical applications. *Radiology.* (1988) **168**:843–9. doi: 10.1148/radiology.168.3.3406414
70. Broche LM, Ross PJ, Davies GR, MacLeod MJ, Lurie DJ. A whole-body Fast Field-Cycling scanner for clinical molecular imaging studies. *Sci Rep.* (2019) **9**:10402. doi: 10.1038/s41598-019-46648-0
71. Basar B, Sonmez M, Paul R, Kocaturk O, Herzka DA, Lederman RJ, et al. Ferromagnetic markers for interventional MRI devices at 0.55 T. In: *Proceedings of the International Society for Magnetic Resonance in Medicine.* Montreal (2019) p. 3845.
72. Bhattacharya I, Ramasawmy R, McGuirt DR, Mancini C, Lederman RJ, Moss J, et al. Improved lung imaging and oxygen enhancement at 0.55 T. In: *Proceedings of the International Society for Magnetic Resonance in Medicine.* Montreal (2019) p. 0003.
73. Restivo MC, Ramasawmy R, Herzka DA, Campbell-Washburn AE. Long TR bSSFP cardiac cine imaging at low field (0.55 T) using EPI and spiral sequences for improved sampling efficiency. In: *Proceedings of the International Society for Magnetic Resonance in Medicine.* Montreal (2019) p. 1084.
74. Ramasawmy R, Herzka DA, Restivo MC, Bhattacharya I, Lederman RJ, Campbell-Washburn AE. Spin-echo imaging at 0.55 T using a spiral trajectory. In: *Proceedings of the International Society for Magnetic Resonance in Medicine.* Montreal (2019). p. 1192.
75. Gries P, Constantinesco A, Brunot B, Facello A. MRI of hand and wrist with a dedicated 0.1-T low-field imaging system. *Magn Reson Imaging.* (1991) **9**:949–53. doi: 10.1016/0730-725X(91)90541-S
76. Lotz H, Ekelund L, Hietala SO, Wickman G. Low field (0.02 T) MRI of the whole body. *J Comput Assist Tomogr.* (1988) **12**:1006–13. doi: 10.1097/00004728-198811000-00018
77. Claudon M, Darrasse L, Boyer B, Regent D, Sauzade M. L'IRM à champ modéré. *Rev Im Med.* (1991) **3**:151–8.
78. Fischer HW, Rinck PA, Van Haverbeke Y, Muller RN. Nuclear relaxation of human brain gray and white matter: analysis of field dependence and implications for MRI. *Magn Reson Med.* (1990) **16**:317–34. doi: 10.1002/mrm.1910160212

Conflict of Interest: MS is a co-founder and former member of Hyperfine Research Inc.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 30 April 2020
doi: 10.3389/fphy.2020.00105

Evaluating the Performance of Ultra-Low-Field MRI for *in-vivo* 3D Current Density Imaging of the Human Head

Peter Hömmen[1†], Antti J. Mäkinen[2*†], Alexander Hunold[3], René Machts[3], Jens Haueisen[3], Koos C. J. Zevenhoven[2], Risto J. Ilmoniemi[2] and Rainer Körber[1]*

[1] Physikalisch-Technische Bundesanstalt, Berlin, Germany, [2] Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, Espoo, Finland, [3] Institute of Biomedical Engineering and Informatics, Technische Universität Ilmenau, Ilmenau, Germany

Magnetic fields associated with currents flowing in tissue can be measured non-invasively by means of zero-field-encoded ultra-low-field magnetic resonance imaging (ULF MRI) enabling current-density imaging (CDI) and possibly conductivity mapping of human head tissues. Since currents applied to a human are limited by safety regulations and only a small fraction of the current passes through the relatively highly-resistive skull, a sufficient signal-to-noise ratio (SNR) may be difficult to obtain when using this method. In this work, we study the relationship between the image SNR and the SNR of the field reconstructions from zero-field-encoded data. We evaluate these results for two existing ULF-MRI scanners—one ultra-sensitive single-channel system and one whole-head multi-channel system—by simulating sequences necessary for current-density reconstruction. We also derive realistic current-density and magnetic-field estimates from finite-element-method simulations based on a three-compartment head model. We found that existing ULF-MRI systems reach sufficient SNR to detect intra-cranial current distributions with statistical uncertainty below 10%. However, the results also reveal that image artifacts influence the reconstruction quality. Further, our simulations indicate that current-density reconstruction in the scalp requires a resolution <5 mm and demonstrate that the necessary sensitivity coverage can be accomplished by multi-channel devices.

Keywords: ultra-low-field MRI, current-density imaging, zero-field encoding, signal-to-noise ratio, finite-element method, Monte-Carlo simulation, MRI simulation

OPEN ACCESS

Edited by:
Elmar Laistler,
Medical University of Vienna, Austria

Reviewed by:
Per Magnelind,
Los Alamos National Laboratory (DOE), United States
Micah Ledbetter,
Kernel, United States

****Correspondence:***
Peter Hömmen
peter.hoemmen@ptb.de
Antti J. Mäkinen
antti.makinen@aalto.fi

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Medical Physics and Imaging, a section of the journal Frontiers in Physics

Received: *20 January 2020*
Accepted: *20 March 2020*
Published: *30 April 2020*

Citation:
Hömmen P, Mäkinen AJ, Hunold A, Machts R, Haueisen J, Zevenhoven KCJ, Ilmoniemi RJ and Körber R (2020) Evaluating the Performance of Ultra-Low-Field MRI for in-vivo 3D Current Density Imaging of the Human Head. Front. Phys. 8:105. doi: 10.3389/fphy.2020.00105

1. INTRODUCTION

Imaging of current-density distributions, produced by injecting current *in vivo* into the human head, has a variety of possible applications. Three-dimensional conductivity distributions or simplified conductivity models may be extracted from such images. These are required for accurate source estimation in electromagnetic neuroimaging [1, 2]. Further, individual conductivity information is necessary for models used to optimize and plan therapeutic treatments, e.g., in transcranial magnetic stimulation (TMS) [3, 4] and transcranial direct-current stimulation (tDCS) [5]. In addition, the current flow during tDCS may be monitored online, providing direct feedback.

Magnetic resonance imaging (MRI) is affected by local magnetic fields, such as the magnetic field $\boldsymbol{B}_J(\boldsymbol{r})$ associated with a current density $\boldsymbol{J}(\boldsymbol{r})$ at points $\boldsymbol{r}$ in the imaging volume. In particular, if

also the main magnetic field B_0 can be switched on and off during the pulse sequence, it is possible to measure full-tensor information of the effects of $\boldsymbol{B}_J(\boldsymbol{r})$, providing a way to directly estimate $\boldsymbol{J}(\boldsymbol{r})$ [6, 7]. The field switching can be achieved [8, 9] in ultra-low-field (ULF) MRI, where the main field is not produced by a persistent superconducting magnet as in conventional high-field MRI. Zero-field-encoded current density imaging (CDI) using superconducting quantum interference device (SQUID)-based ULF MRI was first proposed by Vesanen et al. [6]. It has recently been demonstrated in phantom measurements and is most promising regarding *in-vivo* implementation [9]. Since current impressed *in vivo* in the human head is limited by safety regulations to the low-mA range [10, 11] and only a small fraction of the current passes the relatively highly-resistive skull [12, 13], a sufficient signal-to-noise ratio (SNR) may be difficult to reach.

The two main factors influencing the SNR in ULF MRI are system noise and the strength of the polarizing field that creates the necessary sample magnetization. Both issues have been addressed in previous setups. However, the ultimate sensitivity combining the lowest noise and the highest polarizing field in a single setup has not been demonstrated. Hömmen et al. used an ultra-sensitive single-channel SQUID system with a noise level of 380 aT/$\sqrt{\text{Hz}}$ for the demonstration of CDI [9]. This noise performance was about 10–20 times better than in commercially available SQUID systems, but the polarizing field of 17 mT was comparatively low. Other groups reported ULF-MRI systems with polarizing fields over 100 mT, using cooled copper-coil setups [14, 15]. Even higher polarizing fields could be reached by means of superconducting polarizing coils as presented by Vesanen et al. [16] and Lehto [17].

A quantitative survey of the necessary SNR for zero-field-encoded CDI with a defined uncertainty is still pending. In this work, we investigate the influence of noise on the quality of the $\boldsymbol{B}_J$ and $\boldsymbol{J}$ reconstructions by analytic approximations and by means of Monte-Carlo simulations. Our results enable the estimation of the required image SNR for a given statistical uncertainty in the field reconstructions. They further provide an intuitive method to assess the performance of a specific system for current-density imaging.

In addition, two existing ULF-MRI setups are examined more closely regarding their performance in a CDI application. The first is the single-channel setup of PTB, Berlin, described by Hömmen et al. [9], which is now equipped with an updated polarization setup specially designed for the shape of the human head. The second setup is a whole-head multi-channel system, a successor of the one described by Vesanen et al. [16], located at Aalto University, Helsinki. The latest version comprises an optimized superconductive polarizing coil [17], an ultra-low-noise amplifier for flexible switching of all MRI fields [8], and newly developed SQUID-sensors specially designed for pulsed-field applications [18].

Realistic $\boldsymbol{B}_J$ and $\boldsymbol{J}$ distributions were derived from finite-element-method (FEM) simulations using a three-compartment head model. Combined with nominal gradient fields and sensitivity parameters of the described setups, the $\boldsymbol{B}_J$ distributions were put into a Bloch equation solver that emulates complete gradient-echo sequences in the time domain.

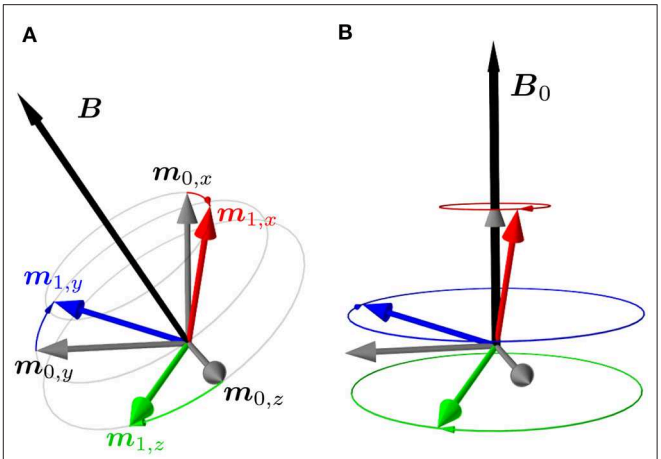


FIGURE 1 | (A) Rotation of three orthogonal starting magnetizations $\boldsymbol{m}_{0,x} = |\boldsymbol{m}_0|\boldsymbol{e}_x$, $\boldsymbol{m}_{0,y} = |\boldsymbol{m}_0|\boldsymbol{e}_y$, and $\boldsymbol{m}_{0,z} = |\boldsymbol{m}_0|\boldsymbol{e}_z$ about the direction of $\boldsymbol{B} = \boldsymbol{B}_\text{B} + \boldsymbol{B}_J$ during the zero-field time. **(B)** After the zero-field time, the main field $\boldsymbol{B}_0 = |\boldsymbol{B}_0|\boldsymbol{e}_x$ ($|\boldsymbol{B}_0| \gg |\boldsymbol{B}|$) is turned on and the vectors start rotating about $\boldsymbol{e}_x$.

Our simulation results not only provide a good estimate of the statistical uncertainty in zero-field-encoded CDI with currently available technologies but also reveal other important requirements in terms of sample coverage and image resolution.

2. ZERO-FIELD-ENCODED CDI

To understand the effects of noise, we recap the sequence and reconstruction method designed by Vesanen et al. [6]. More detailed information on the experimental implementation, including the sequence diagram, can be gleaned from Hömmen et al. [9].

At first, magnetization is built up by a polarization period. Subsequently, all MRI fields are turned off and the current density $\boldsymbol{J}$ is applied during a defined zero-field time τ. After the zero-field time, the magnetization has been rotated to $\boldsymbol{m}_1$ by the magnetic field during τ as

$$\boldsymbol{m}_1(\boldsymbol{r}) = e^{\gamma\tau\mathbf{A}(\boldsymbol{r})}\boldsymbol{m}_0(\boldsymbol{r}) = \Phi(\boldsymbol{r})\boldsymbol{m}_0(\boldsymbol{r})\,, \tag{1}$$

where $\boldsymbol{m}_0$ is the starting magnetization and γ is the gyromagnetic ratio of the proton. $\mathbf{A}$ is the generator of the rotation matrix $\boldsymbol{\Phi}$, which describes the magnetization dynamics due to the quasi-static magnetic field during τ [19, p. 86–89] [6]. Ideally, this field is solely determined by the magnetic field $\boldsymbol{B}_J$ associated with $\boldsymbol{J}$. In reality, a superposition of a static background field and transient fields due to pulsing (in the following combined in the term $\boldsymbol{B}_\text{B}$) are present. Hence, the time evolution of $\boldsymbol{m}$ is affected by $\mathbf{A}_J + \mathbf{A}_\text{B}$, where $\mathbf{A}_J$ and $\mathbf{A}_\text{B}$ are associated with $\boldsymbol{B}_J$ and the average $\boldsymbol{B}_\text{B}$, respectively.

Following τ, the main field $\boldsymbol{B}_0$, here in the x-direction, is turned on and the magnetization is manipulated by gradient fields to encode spatial information in the phase and frequency

of the resulting signal. Ignoring relaxation, the magnetic signal recorded at a sensor during the echo can be written as

$$\begin{aligned} S(t) &= \int \boldsymbol{C}(\boldsymbol{r})^\top \boldsymbol{m}(\boldsymbol{r},t)dV \\ &= \int \boldsymbol{C}(\boldsymbol{r})^\top \mathbf{R}_\mathrm{f}(\boldsymbol{r},t)\mathbf{R}_\mathrm{p}(\boldsymbol{r})\boldsymbol{m}_1(\boldsymbol{r})dV\,, \end{aligned} \tag{2}$$

where t is the time, $\boldsymbol{C}$ the coupling field of the sensor, and matrices $\mathbf{R}_\mathrm{f}$ and $\mathbf{R}_\mathrm{p}$ correspond to rotations in the yz plane during the frequency- and phase-encoding periods. For the following operations, it is convenient to convert the signal equation to a complex representation. Considering only the frequency components close to the Larmor angular frequency $\gamma|\boldsymbol{B}_0|$, the signal can be written as [20, 21]

$$S(t) \approx \mathrm{Re}\int \beta(\boldsymbol{r})^* e^{i[\omega(\boldsymbol{r})t+\theta_\mathrm{p}(\boldsymbol{r})]}\tilde{m}_1(\boldsymbol{r})dV\,, \tag{3}$$

where $\beta = C_z + iC_y$, $\tilde{m}_1 = m_{1,z} + im_{1,y}$, ωt is the phase angle due to precession during frequency encoding, and θ_p the angle due to phase encoding. In a realistic setting, β could also include additional effects from an inhomogeneous polarizing field and non-idealities in field pulsing.

After applying the discrete Fourier transform to the frequency- and phase-encoded data and taking the relevant frequency bins, the magnitude and phase of the rotation of $\boldsymbol{m}$ can be estimated at the location of the corresponding voxel. The voxel value corresponding to the MR signal generated close to $\boldsymbol{r}_n$ is given by

$$\begin{aligned} \nu_n &= \int \mathrm{SRF}(\boldsymbol{r}-\boldsymbol{r}_n)\beta(\boldsymbol{r})^*\tilde{m}_1(\boldsymbol{r})dV \\ &\approx \beta^*(\boldsymbol{r}_n)\tilde{m}_1(\boldsymbol{r}_n)\,, \end{aligned} \tag{4}$$

where $\mathrm{SRF}(\boldsymbol{r}-\boldsymbol{r}_n)$ is the spatial response function of the n^th voxel [22]. When the SRF is close to a delta function $\delta(\boldsymbol{r}-\boldsymbol{r}_n)$, the integral can be approximated with the function value at $\boldsymbol{r}_n$, otherwise the SRF will result in leakage artifacts from the neighboring areas.

The voxel values ν_n contain information about the zero-field-encoded magnetic field in both their magnitude and phase. In reality, there are other factors, such as non-idealities in the gradient ramps and unknown relaxation profiles, that affect the voxel values as well. Therefore, the relative changes in ν_n associated with the current density are recovered by normalization with a reference u_n [6, 9]. Repeating the sequence for three orthogonal starting magnetizations $\boldsymbol{m}_{0,x} = |\boldsymbol{m}_0|\boldsymbol{e}_x$, $\boldsymbol{m}_{0,y} = |\boldsymbol{m}_0|\boldsymbol{e}_y$ and $\boldsymbol{m}_{0,z} = |\boldsymbol{m}_0|\boldsymbol{e}_z$ (shown in **Figure 1**), the last two rows of $\boldsymbol{\Phi}_n$ can be measured. For example, the y and z elements of the first column are given by:

$$\begin{aligned} \Phi_{n(31)} &= \mathrm{Re}[\nu_{n,x}/u_n] \\ \Phi_{n(21)} &= \mathrm{Im}[\nu_{n,x}/u_n]\,, \end{aligned} \tag{5}$$

where $\nu_{n,x}$ denotes the voxel value of a zero-field-encoded image with starting magnetization in the x direction. Rotation matrices are orthogonal by definition. Therefore, the first row of $\boldsymbol{\Phi}_n$ can be derived by the cross product of the second with the third row. Naturally, the elements in $\boldsymbol{\Phi}_n$ are contaminated by noise. A practical approach to increase the accuracy is to apply an orthogonalization. For this purpose Vesanen et al. [6] suggest Löwdin's transformation, which yields the closest orthogonalization in the least-squares sense [23, 24]. It is clear that a unique rotation matrix $\boldsymbol{\Phi}_n$ is created for each voxel n. The following analysis in this section and in section 3 concentrates on a voxel-wise reconstruction of $\boldsymbol{B}_J$ and $\boldsymbol{J}$, where the index n is left out for simplicity.

Using $\boldsymbol{\Phi}$, all components of the magnetic field $\boldsymbol{B} = \boldsymbol{B}_\mathrm{B} + \boldsymbol{B}_J$ can be derived from a non-linear inversion of the matrix exponential:

$$\begin{aligned} \gamma\tau\mathbf{A} &= \gamma\tau\begin{bmatrix} 0 & \hat{B}_z & -\hat{B}_y \\ -\hat{B}_z & 0 & \hat{B}_x \\ \hat{B}_y & -\hat{B}_x & 0 \end{bmatrix} \\ &= \frac{\phi}{2\sin\phi}(\boldsymbol{\Phi}-\boldsymbol{\Phi}^\top)\,, \end{aligned} \tag{6}$$

where $\phi = \arccos[(\mathrm{tr}(\boldsymbol{\Phi})-1)/2]$ represents the rotation angle of $\boldsymbol{\Phi}$ [6], and $\hat{\boldsymbol{B}}$ is the reconstruction of $\boldsymbol{B}$. From here on, reconstructed quantities are denoted using the hat symbol.

Finally, $\boldsymbol{B}_J$ can be estimated by subtracting another reconstruction from $\hat{\boldsymbol{B}}$. This could be a full 3D image of $\boldsymbol{B}_\mathrm{B}$ only, or of $\boldsymbol{B}_\mathrm{B} + \boldsymbol{B}_J$ with the impressed current having the opposite polarity. The latter reduces the statistical uncertainty by $1/\sqrt{2}$ and is from here on called bipolar reconstruction:

$$\begin{aligned} \hat{\boldsymbol{B}}_J &= \frac{\hat{\boldsymbol{B}}_1 - \hat{\boldsymbol{B}}_2}{2}\,, \\ \hat{\boldsymbol{B}}_1 &= \hat{\boldsymbol{B}}_\mathrm{B} + \hat{\boldsymbol{B}}_{J(+)}\,, \\ \hat{\boldsymbol{B}}_2 &= \hat{\boldsymbol{B}}_\mathrm{B} + \hat{\boldsymbol{B}}_{J(-)}\,. \end{aligned} \tag{7}$$

From Equation (7), the full tensor of the local field $\hat{\boldsymbol{B}}_J$ is derived, enabling the estimation of $\hat{\boldsymbol{J}}$ by Ampère's Law:

$$\hat{\boldsymbol{J}} = \frac{1}{\mu_0}\nabla\times\hat{\boldsymbol{B}}_J\,, \tag{8}$$

where μ_0 is the permeability of free space.

3. NOISE IN ZERO-FIELD CDI

3.1. The Connection Between Noise in $\boldsymbol{\Phi}$ and Image SNR

In this section, we analyze how the uncertainty in the reconstruction of zero-field-encoded data relates to the image SNR. From Equation (5), we know that the values in $\boldsymbol{\Phi}$ are normalized by a complex reference $u = |u|e^{i\delta}$, where $|u|$ is related to the magnitude of the magnetization after τ and δ to the phase accumulation due to effects that do not arise from $\boldsymbol{B}_J + \boldsymbol{B}_\mathrm{B}$.

Hömmen et al. [9] describe that $|u|$ cannot be measured directly due to the always present background field. However, the reference can be constructed from the real or imaginary parts of the three measurements of v by

$$|u| = \sqrt{\mathrm{Re}[v_x]^2 + \mathrm{Re}[v_y]^2 + \mathrm{Re}[v_z]^2}\,, \tag{9}$$

which effectively normalizes the rows of $\mathbf{\Phi}$ to exactly unit norm. The reference phase δ, on the other hand, has to be acquired in a separate measurement. See Hömmen et al. [9] for more detail.

The complex reference value can be modeled as $u = \mathrm{E}[u] + \epsilon$, where E denotes the expected value and $\epsilon \sim \mathcal{N}(0, \sigma^2)$ is symmetric complex Gaussian noise that can be extracted from a noise-only image e, or from a noise-only region in any of the images v. Using this reference, we define the image SNR as

$$\begin{aligned}\mathrm{SNR} &\overset{\mathrm{def}}{=} \frac{|\,\mathrm{E}[u]|}{\mathrm{SD}[e]} = \frac{|\,\mathrm{E}[u]|}{\sqrt{\mathrm{E}[\mathrm{Re}(\epsilon)^2] + \mathrm{E}[\mathrm{Im}(\epsilon)^2]}} \\ &= \frac{|\,\mathrm{E}[u]|}{\sigma}\,,\end{aligned} \tag{10}$$

where SD is the standard deviation.

The phase correction with the noisy reference phase δ causes the real part to leak to the imaginary part and vice versa, increasing the noise in the matrix elements. Dividing by the magnitude of the complex reference $u = |u|e^{i\delta}$ yields unit norm in the rows of $\mathbf{\Phi}$ decreasing the noise. This is derived in the **Appendix**, which also shows that the noise SD in the elements of $\mathbf{\Phi}$ can be approximated as

$$\sigma_{\Phi_{ij}} = \frac{1}{\sqrt{2}\,\mathrm{SNR}} g_{ij}(\mathbf{\Phi})\,, \tag{11}$$

where the scaling $1 \leq g_{ij}(\mathbf{\Phi}) \leq \sqrt{2}$ depends on the associated measurement. This approximation is valid when $u \approx \mathrm{E}[u]$, i.e., $\mathrm{SNR} \gg 1$. Equation (11) already gives an impression of the noise SD in $\mathbf{\Phi}$ as a function of the image SNR. The rotation-dependent scaling $g_{ij}(\mathbf{\Phi})$ and correlations between the elements are given in the **Appendix**.

The most important factors determining the SNR are the polarizing field, the coupling to the sensors, and the relaxation of the magnetization, all of which affect the voxel magnitude. The noise in the voxel values is governed by the system noise determined by the magnetic sensor as well as other instrumentational and environmental noise sources.

3.2. Noise Analysis of B-Field Reconstruction: Linear Approximation

To estimate the noise in the reconstruction of $\boldsymbol{B}$, we first discuss an idealized case, where all three rows of $\mathbf{\Phi}$ can be measured and no reference image u is needed. In this case, the noise in the elements of $\mathbf{\Phi}$ becomes independent and identically distributed with standard deviation of $1/(\sqrt{2}\,\mathrm{SNR})$.

We start by using a first-order small-angle approximation of the rotation matrix

$$\mathbf{\Phi} \approx \mathbf{I} + \gamma\tau\mathbf{A} = \begin{bmatrix} 1 & \gamma\tau B_z & -\gamma\tau B_y \\ -\gamma\tau B_z & 1 & \gamma\tau B_x \\ \gamma\tau B_y & -\gamma\tau B_x & 1 \end{bmatrix}, \tag{12}$$

where $\mathbf{I}$ is the identity matrix. The magnetic field components can be solved directly and, as each component is measured twice, they can be averaged so that the noise SD in the angular quantity becomes $\sigma_{\gamma\tau\hat{B}_d} = 1/(2\,\mathrm{SNR})$. Here, d is any of the components x, y or z, and the noise SD of a magnetic field component can be derived to $\sigma_{\hat{B}_d} = 1/(2\gamma\tau\ \mathrm{SNR})$.

In reality, the elements of $\mathbf{\Phi}$ are estimated with the help of a reference image, which modifies the noise in the elements as derived in the **Appendix**. Additionally, only two rows of the rotation matrix $\mathbf{\Phi}$ can be obtained from the measurements as explained in section 2. Therefore, one row (in our case the first row) has to be derived from the cross product of the adjacent rows, where the cross product contains information about the components of $\boldsymbol{B}$ orthogonal to the direction of $\boldsymbol{B}_0$. These components are no longer subject to independent random noise; consequently, the noise is not reduced by the averaging effect in the linear reconstruction.

So far, the noise analysis was discussed for the reconstruction of the effective $\boldsymbol{B}$-field. As mentioned before, in practice, the measurement of $\boldsymbol{B}_J$ is contaminated by a background field $\boldsymbol{B}_\mathrm{B}$, which must be eliminated by subtracting a second reconstruction. The noise in the two reconstructions is independent, which is why in the case of bipolar reconstruction the noise in the field estimate is reduced by a factor of $\sqrt{2}$ (see Equation 7). Additionally, as the reference phase δ is the same for the two data sets, the additional noise due to referencing will cancel in the field subtraction.

In the first-order approximation, we finally obtain for bipolar reconstruction

$$\sigma_{\hat{B}_y} = \sigma_{\hat{B}_z} \approx \frac{1}{2\gamma\tau\ \mathrm{SNR}} \tag{13}$$

and

$$\sigma_{\hat{B}_x} \approx \frac{1}{2\sqrt{2}\gamma\tau\ \mathrm{SNR}}\,, \tag{14}$$

because B_x is measured twice.

3.3. Noise Analysis of B-Field Reconstruction: Monte-Carlo Simulations

From the first-order small-angle approximation we can gain intuitive understanding of the statistical uncertainty in the reconstruction of $\boldsymbol{B}_J$. However, in reality, the rotation angle ϕ can obtain values up to π and the linear approximation breaks down.

In order to estimate the influence of noise on the non-linear reconstruction, we carried out a series of Monte-Carlo simulations. Therefore, we generated the last two rows of rotation matrices $\mathbf{\Phi}$ for 100 different rotation angles $\phi = \pm\gamma\tau|\boldsymbol{B}|$ taken uniformly between $-\pi < \phi < \pi$, where the negative angles correspond to $-\boldsymbol{B}$. As before, $\boldsymbol{B} = \boldsymbol{B}_\mathrm{B} + \boldsymbol{B}_J$, where $\boldsymbol{B}_J$ was set to zero and ϕ was varied by adjusting $\boldsymbol{B}_\mathrm{B}$. The matrices $\mathbf{\Phi}$ were generated using the general formula of Rodriguez, as explained in [19, p. 86–89]:

$$\mathbf{\Phi} = e^{\phi\mathbf{K}} = \mathbf{I} + \sin(\phi)\mathbf{K} + (1 - \cos(\phi))\mathbf{K}^2\,. \tag{15}$$

Here, $\mathbf{K} = \gamma\tau\mathbf{A}/\phi$ is a unitary cross-product matrix associated with the rotation axis. Independent and Gaussian-distributed random noise was generated and superimposed with each element of $\mathbf{\Phi}$, according to Equation (11). Subsequently, the first row was derived by the cross product of the other two. The procedure was repeated 100,000 times to obtain statistics for the reconstruction quality.

Figure 2 illustrates the standard deviation after three intermediate steps of the reconstruction, showcasing their influences on the result. The data are normalized to the input noise $1/(\sqrt{2}\,\mathrm{SNR})$ corresponding to Equation (11) without $g_{ij}(\mathbf{\Phi})$.

Figure 2A illustrates a case where no referencing with u was applied. Each element of $\mathbf{\Phi}$ thus contained the same amount of Gaussian distributed noise. Although this may not be the case in an experimental implementation, one sees that $\hat{B}_x$ contains $1/\sqrt{2}$ the noise of the other components for small angles of ϕ, as predicted by the first-order approximation. However, with a rising field strength, i.e., larger rotation angle ϕ, the noise in this component increases non-linearly and more strongly compared to the components orthogonal to $\boldsymbol{B}_0$.

The simulations underlying **Figure 2B** include the necessary pre-referencing. For very small angles, the extra phase noise due to the noisy reference phase δ affects the noise SD only in $\hat{B}_x$. Toward larger angles, this effect is visible in $\hat{B}_z$. The y-component of $\hat{\boldsymbol{B}}$ is not affected, which is in accordance with the analysis presented in the **Appendix**.

Figure 2C shows the results after subsequent orthogonalization using the Löwdin transformation. We observe a strong effect toward large angles ϕ, especially in the x-component, which is parallel to $\boldsymbol{B}_0$.

Figure 3 illustrates the standard deviations of the results of a simulated bipolar reconstruction. In comparison to **Figure 2**, these data sets are arithmetic means of two similar fields (independent noise, identical reference), respectively Equation (7) with $\boldsymbol{B}_J = 0$. Overall, the noise levels decrease by a factor of $\sqrt{2}$, in comparison to the reconstructions of the effective field $\boldsymbol{B}$ in **Figure 2**. Further, the additional noise due to the reference phase δ, visible in **Figures 2B,C**, was subtracted entirely. Except for very large angles ($\phi > 7\pi/8$), the noise SD in each component is lower than $1/(\mathrm{SNR}\sqrt{2})$. **Figure 3** also shows a measure to assess the expected deviation from the mean of $\hat{\boldsymbol{B}}_J$ (purple line), which can be derived to be the square root of the trace of the covariance matrix:

$$\begin{aligned}\mathrm{SD}[\hat{\boldsymbol{B}}_J] &= \sqrt{\mathrm{E}\left[|\hat{\boldsymbol{B}}_J - \mathrm{E}(\hat{\boldsymbol{B}}_J)|^2\right]} \\ &= \sqrt{\mathrm{tr}\left[\mathrm{cov}(\hat{\boldsymbol{B}}_J)\right]} \\ &= \sqrt{\sigma^2_{\hat{B}_{J,x}} + \sigma^2_{\hat{B}_{J,y}} + \sigma^2_{\hat{B}_{J,z}}}\,.\end{aligned} \tag{16}$$

3.4. Noise Analysis of Current-Density Reconstruction

From the noise in the reconstruction of $\boldsymbol{B}_J$, we can also calculate the noise in the current density reconstruction using Equation (8). For that, we make some simplifications. We assume a constant current density in a homogeneous and isotropic medium. Further, we assume a homogeneous background field that is much larger than $\boldsymbol{B}_J$. A simple method for the spatial derivation is to take into account only the two nearest neighbors at $z - l$ and $z + l$

$$\frac{d\hat{\boldsymbol{B}}_J}{dz}(z) = \frac{\hat{\boldsymbol{B}}_J(z+l) - \hat{\boldsymbol{B}}_J(z-l)}{2l}\,, \tag{17}$$

where z is the coordinate of the voxel in the z-direction and l is the voxel sidelength. Assuming equal SNR at $z + l$ and $z - l$, the noise SD of the gradient is approximately $\sigma_{\boldsymbol{G}(z)} = \sigma_{\hat{\boldsymbol{B}}_J(z)}/(l\sqrt{2})$. Applying the curl

$$\hat{J}_x = \frac{1}{\mu_0}\left(d\hat{B}_{J,z}/dy - d\hat{B}_{J,y}/dz\right) \tag{18}$$

and neglecting the small possible differences in $\sigma_{\hat{B}_{J,z}}$ and $\sigma_{\hat{B}_{J,y}}$, the noise SD of $\hat{J}_x$ can be approximated as $\sigma_{\hat{J}_x} = \sigma_{\hat{B}_{J,z}}/(l\mu_0)$.

3.5. Field Reconstruction Quality in Terms of Image SNR

Using the definition of image SNR in Equation (10) and the results of the Monte-Carlo simulations, the signal-to-noise ratio of the $\boldsymbol{B}_J$ reconstruction ($\mathrm{SNR}[\hat{\boldsymbol{B}}_J]$) can be estimated by

$$\begin{aligned}\mathrm{SNR}[\hat{\boldsymbol{B}}_J] &\overset{\mathrm{def}}{=} \frac{|\hat{\boldsymbol{B}}_J|}{\mathrm{SD}[\hat{\boldsymbol{B}}_J]} \\ &= \frac{\gamma\tau|\hat{\boldsymbol{B}}_J|\sqrt{2}}{c}\mathrm{SNR}\,,\end{aligned} \tag{19}$$

where $\mathrm{SD}[\hat{\boldsymbol{B}}_J]$ is the measure for noise in the vector $\hat{\boldsymbol{B}}_J$ defined in Equation (16). Further, the scaling factor c depends on the strength and the orientation of $\boldsymbol{B}_\mathrm{B}$ and can be read directly from the purple, dash/dotted lines in **Figure 3**. As c is highest for x-directional background fields, a polynomial, normalized to $1/\pi$, was fitted to the data presented in **Figure 3B**, to approximate c as a function of ϕ:

$$c(\phi) \approx 0.17\left(\frac{\phi}{\pi}\right)^4 + 0.35\left(\frac{\phi}{\pi}\right)^2 + 1.118\,. \tag{20}$$

Note that the results presented in **Figures 3A,C** only deviate slightly from Equation (20).

According to the figure, without any information on the background field, a representative value for the scaling factor would be $c = 1.3$. This is close to the worst-case scenario as higher rotation angles may cause phase wrapping.

To provide a numerical example, let us assume that $|\boldsymbol{B}_J| = 10$ nT, a homogeneous x-directional background field of 60 nT, and a zero-field time of $\tau = 100$ ms, taking into account the T_2-relaxation time of gray matter in the μT regime of approximately 100 ms. Substituting the rotation angle $\phi = \gamma\tau|\boldsymbol{B}|$ in Equation (20), c is approximated to be 1.2. According to Equation (19), for a required $\mathrm{SNR}[\hat{\boldsymbol{B}}_J] > 10$, the voxel SNR needs to be over 32.

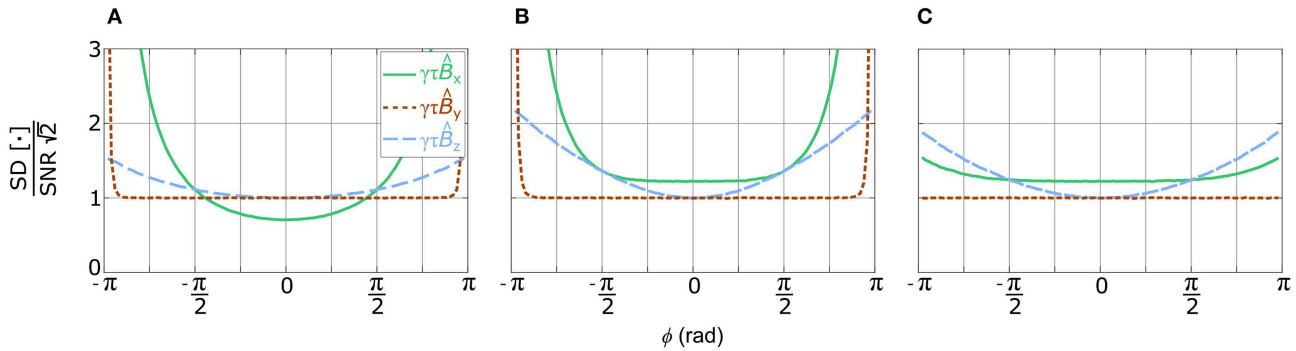


FIGURE 2 | Single-voxel Monte-Carlo simulations to estimate the influence of noise on three different steps of the non-linear reconstruction as a function of the rotation angle ϕ. The shown data are based on simulated noisy rotation matrices, where the first row was derived by the cross product of the other two. Displayed are normalized standard deviations of each component of $\hat{\boldsymbol{B}}$, which is the reconstruction of y-directional field $\boldsymbol{B} = |\boldsymbol{B}_\mathrm{B}|\boldsymbol{e}_y$. $|\boldsymbol{B}_\mathrm{B}|$ was adjusted to generate the rotation angles ϕ with the negative angles corresponding to the field direction $-\boldsymbol{e}_y$. The main field $\boldsymbol{B}_0$ was x-directional. The figures show the standard deviations of reconstructions without pre-referencing **(A)**, with pre-referencing **(B)**, and with subsequent orthogonalization using Löwdin's transformation **(C)**.

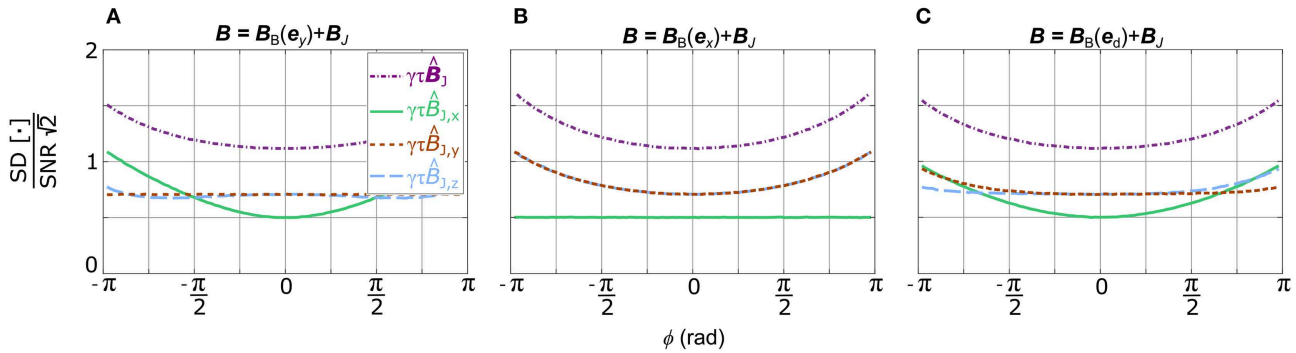


FIGURE 3 | Single-voxel Monte-Carlo simulations to estimate the standard deviation of each component of $\hat{\boldsymbol{B}}_J$ after bipolar reconstruction (Equation 7), in dependence of the rotation angle ϕ. In addition, $\sqrt{\mathrm{tr}[\mathrm{cov}(\hat{\boldsymbol{B}}_J)]}$ (Equation 16) is presented in purple, dash/dotted lines. $\boldsymbol{B}$ is the effective field $\boldsymbol{B}_\mathrm{B} + \boldsymbol{B}_J$, where $\boldsymbol{B}_J$ was set to zero and $\boldsymbol{B}_\mathrm{B}$ was adjusted to generate defined rotation angles ϕ with negative angles corresponding to $-\boldsymbol{B}$. The figures represent reconstructions, where $\boldsymbol{B}_\mathrm{B}$ was y-directional **(A)**, x-directional **(B)**, and diagonally oriented in $\boldsymbol{e}_\mathrm{d} = [1, 1, 1]/\sqrt{3}$ **(C)**. The main field $\boldsymbol{B}_0$ was x-directional in all cases.

The estimation of $\boldsymbol{J}$ using Ampère's law requires the determination of local field gradients, where the noise in the reconstruction is inversely proportional to the voxel side length l. This effect should not be underestimated; as the signal strength already scales to the voxel volume l^3, the SNR of $\hat{\boldsymbol{J}}$ scales to the fourth power of the voxel sidelength. The quality of the $\boldsymbol{J}$-reconstruction can be determined from the SNR of $\hat{\boldsymbol{B}}_J$, by including the scaling factor $l\mu_0$ in Equation (19):

$$\begin{aligned}\mathrm{SNR}[\hat{\boldsymbol{J}}] &\overset{\mathrm{def}}{=} \frac{|\hat{\boldsymbol{J}}|}{\mathrm{SD}\,[\hat{\boldsymbol{J}}]} \\ &\approx \frac{\gamma\tau l\mu_0|\hat{\boldsymbol{J}}|\sqrt{2}}{c}\mathrm{SNR}\,.\end{aligned} \tag{21}$$

The approximation in Equation (21) is valid when the voxels involved in the gradient estimation are subject to equal SNR. Especially at tissue boundaries, this can cause erroneous assessments due to different relaxation times.

Again, to provide an example, we assume a current density distribution of 0.4 A/m^2, a value in accordance with the literature for a stimulation of approximately 4 mA [13]. Similar to the example above, $c \approx 1.2$ is assumed. If we want to derive $\hat{\boldsymbol{J}}$ with $\mathrm{SNR}[\hat{\boldsymbol{J}}] > 10$ and a voxel-sidelength of 5 mm, a required image SNR of 130 is estimated.

4. SIMULATED PERFORMANCE OF ULF-MRI SYSTEMS

4.1. MRI Simulation Setup

The main factors that determine the SNR profiles of ULF-MR images are the sensor arrangement, system noise, and the polarizing field profile. To evaluate the sensitivity of the $\hat{\boldsymbol{B}}_J$ and $\hat{\boldsymbol{J}}$ field reconstruction in a realistic situation, we set up a simulation toolbox incorporating realistic polarizing fields and sensor geometries, as well as time-domain magnetization evolution based on analytical solutions of Bloch's equation. Assuming ideal gradient fields and instantaneous field switching,

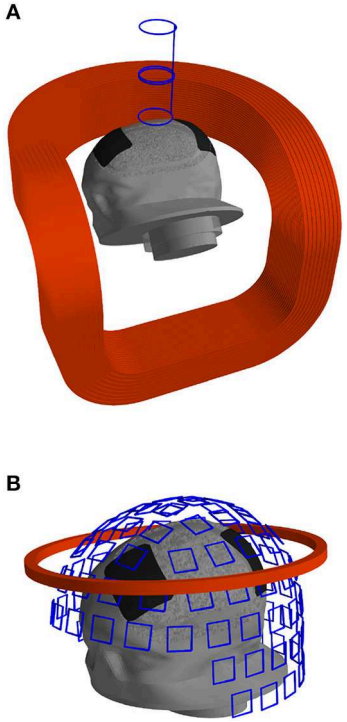


FIGURE 4 | Geometries of the single-channel MRI setup at PTB **(A)** and the multi-channel MRI setup at Aalto **(B)**. The illustrations include the polarizing coil (red), the receiver coils of the sensors (blue), the head model (gray), and the stimulation electrodes (black).

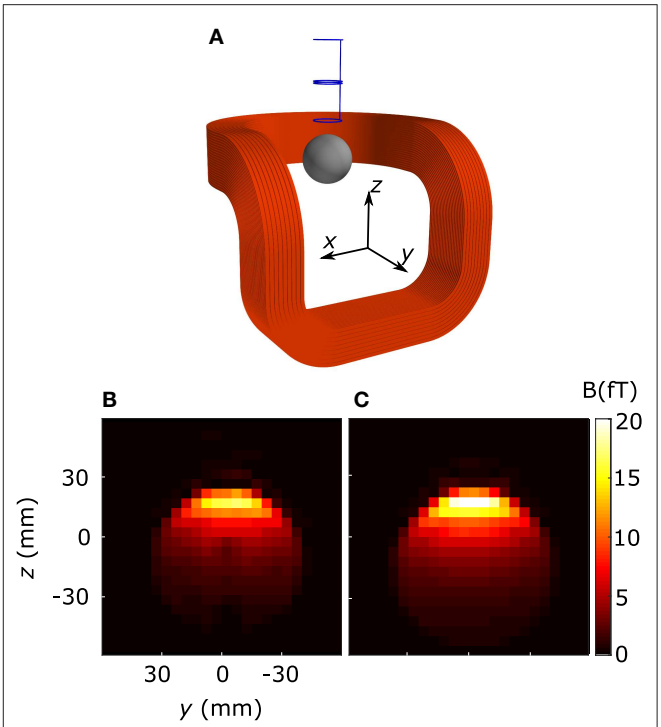


FIGURE 5 | Comparison of measured and simulated MRI images. **(A)** shows the utilized setup, including the polarizing coil (red), the spherical phantom (gray), and the receiver coil of the sensor (blue). Central slices of reconstructed images, not corrected for the sensitivity profile, are presented for measurement **(B)** and simulation **(C)**. Please note that the actual phantom contains a mount for dipolar current electrodes, which is recognizable in the central lower half of the reconstructed measurement, but was not accounted for in the simulations.

gradient-echo sequences can be simulated for arbitrary imaging objects. Both the polarizing field profile and the coupling of the magnetization to the sensor (Equation 2) were calculated by analytically integrating the Biot–Savart formula over line segments [20, 25].

Two sets of simulations were set up to correspond to the single-channel system with a wire-wound 2nd-order axial gradiometer and a resistive polarizing coil as present at PTB, Berlin, and the multi-channel whole-head system with 102 planar thin-film magnetometers and a compact superconducting polarizing coil built at Aalto University (see **Figure 4**). Based on measured values, the sensor noise in the single-channel system was set to 350 aT/$\sqrt{\text{Hz}}$ and in the multi-channel system to 2 fT/$\sqrt{\text{Hz}}$. A polarizing current of 50 A was chosen for both setups corresponding to field maximum of 90 mT and mean of 65 mT in the brain compartment for the single-channel system. For the multi-channel system, the field maximum was 115 mT and the mean 70 mT in the brain compartment.

For the evaluation of the simulations, a comparison with actual measurements using the PTB setup was executed. Therefore, a spherical single-compartment phantom (80 mm diameter), filled with an aqueous solution of $CuSO_4$+H_2O to tune the T_2-relaxation time to approximately 100 ms, was placed with a gap of 10 mm below the dewar (nominal warm-cold distance 13 mm). The current in the polarizing coil was set to 20 A, resulting in an inhomogeneous polarizing field of approximately 25 mT. Gradients were set to give a voxel size of (4.8 × 4.8 × 4.8) mm^3 and a field of view (FOV) of 115 mm in the phase-encoded directions y and z. The resulting time signals of the gradient echos were processed to form an array of k-space data. To reduce Gibbs ringing, both the frequency- and the phase-encoding dimensions were tapered with a Tukey window (shape parameter = 0.5) and the three-dimensional FFT was applied to reconstruct the images. For the simulations, the sphere was approximated by a regular 1-mm spaced grid.

Figure 5 illustrates the setup, accompanied by magnitude images of measurement and simulation. The results reveal a difference in the amplitude of measured and simulated MRI of approximately 25%, probably subject to multiple origins. A shielding coil reduces the polarizing field of the actual setup, which was not accounted for in the simulations. Also, winding errors due to the relatively complex geometry of the polarizing coil reduce the current–field ratio. In addition, the true warm–cold distance of the dewar could vary depending on the helium level and the phantom mount also might have inaccuracy in the mm range. Taking all these uncertainties into account, the simulated MRI sequence resembles the realistic conditions found in actual measurements.

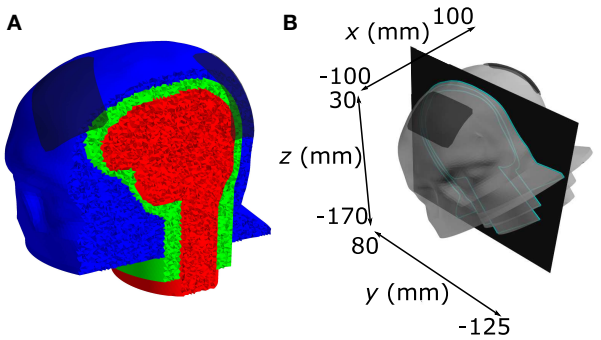


FIGURE 6 | (A) Tetrahedral FEM mesh consisting of intra-cranial volume (red), skull (green), and scalp (blue) compartments. The electrodes are illustrated in transparent gray. **(B)** Model positioning in the MRI coordinate system. The plane between the electrodes corresponds to the slice in **Figures 8**, **9**.

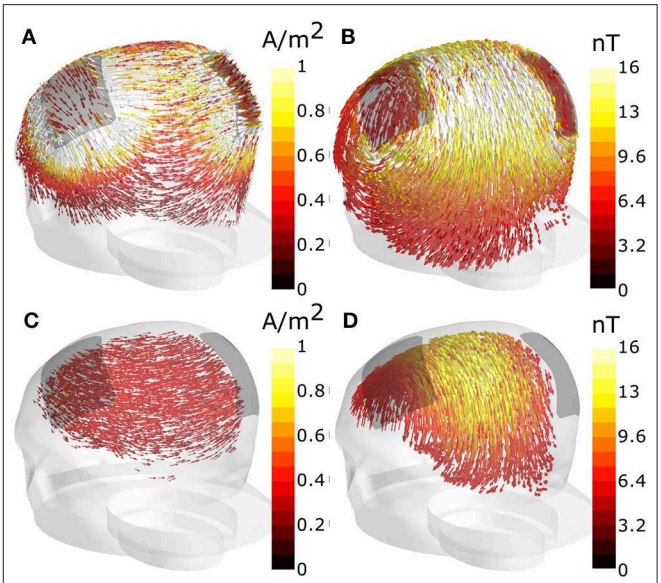


FIGURE 7 | The FEM simulation results for current density $\boldsymbol{J}$ are visualized in the scalp **(A)** and in the brain compartment **(C)**. The simulated magnetic field $\boldsymbol{B}_J$, due to all current flowing in the head, is plotted in the scalp **(B)** and in the brain **(D)**. The arrow lengths are scaled logarithmically because of the vast magnitude differences especially in the current density. Each subfigure shows only the top 30 (magnitude) percentile of the field in the respective compartment.

4.2. MRI Simulations With Head Model

In the next step, the simulation setup was used to generate full CDI sequences with the single-channel system, as well as with the multi-channel system, using the $\boldsymbol{B}_J$ distribution derived from finite-element-method (FEM) simulations of a realistic head model. This model is based on CT scans of a human head [26] and contains three compartments as shown in **Figure 6A**. The conductivity in the outermost scalp compartment was set to 0.22 S/m, in the skull compartment to 0.01 S/m, and in the innermost brain compartment to 0.33 S/m. The two stimulation electrodes were positioned roughly 10 cm apart, one on the forehead and the other one on the side of the head. The electrode dimensions were (50 × 70) mm^2 and their conductivity was set to 1.4 S/m.

The FEM simulations to obtain the current density $\boldsymbol{J}$ and the resulting magnetic field $\boldsymbol{B}_J$ were conducted in the Comsol Multiphysics software based on the generalized minimal residual method (GMRES). Current flow was realized by setting zero potential on the outer surface of the cathode and applying a total current of 4.5 mA to the outer surface of the anode. For the calculation of $\boldsymbol{B}_J$, a spherical air compartment (2 m in diameter) was added to the model, ensuring a negligible effect of the magnetic isolation boundary condition.

For the MRI simulations, the head model was positioned in the FOV of the two described scanner arrangements, similar to how the positioning of a head would be in an actual measurement setup (compare with **Figure 4**). The scalp–sensor distance was 16 mm for the single-channel setup and 20–35 mm for the multi-channel setup, taking into account the individual warm–cold distances of the two systems plus 3 mm to compensate for the amplitude differences found in the comparison with actual measurements, as described in section 4.1. The magnetization was discretized to tetrahedral elements derived from the geometry of the Comsol model. The time evolution of the magnetic moment was simulated for the center of each element. The T_2-relaxation time for the brain compartment was set to 106 ms and for the scalp compartment to 120 ms [27]. For simplicity, as the spin density in the skull is insignificant compared to soft tissue, this compartment was assumed to have no magnetization at all. The average tetrahedron sidelengths were approximately 3.5 mm in the brain and 2.5 mm in the scalp. Gradients were set to give a voxel size of (5 × 5 × 5) mm^3 and a field of view (FOV) of 220 mm in the phase-encoded directions. **Figure 6B** presents the head model in the coordinate system defined by the MRI gradients. As performed for the spherical phantom, both the frequency- and phase-encoding dimensions were tapered with a Tukey window (shape parameter = 0.5) before computing the three-dimensional FFT. For the multi-channel system, images of each sensor were combined voxel-wise using the coupling field information as described in Zevenhoven et al. [20].

4.3. Simulation Results

Patterns of the simulated current density and the associated magnetic field, as derived from the FEM simulations, are shown in **Figure 7**. Due to the low conductivity of the skull, the highest current density can be found in the scalp compartment. In the vicinity of the electrode boundary, $|\boldsymbol{J}|$ was up to 15 A/m^2. The maximal current density in the brain compartment below the electrodes was about 0.5 A/m^2. In relation to that, the magnetic field appeared smoother, yielding maximal field strengths of 20 nT in the scalp and 12 nT in the brain compartment. The maximum of the field magnitude in the brain compartment is located in between the electrodes, just beneath skull layer. In contrast, the maximal current density in the brain is located beneath the electrodes.

Figure 8 shows a comparison between the field reconstructions $|\hat{\boldsymbol{B}}_J|$ and $|\hat{\boldsymbol{J}}|$ of the simulated zero-field sequence and ground-truth FEM solutions of $|\boldsymbol{B}_J|$ and $|\boldsymbol{J}|$. Both data sets are presented without noise. The plane corresponding to the slice is defined in **Figure 6B**. The reconstructed magnetic field $|\hat{\boldsymbol{B}}_J|$ resembles closely the corresponding FEM solution, which was used as an input to the MR simulations. Notable differences are found inside the skull, which is expected due to the lack of magnetization, as well as on the top parts of the scalp at the field maximum. The difference image reveals ringing artifacts in the intra-cranial volume, leading to error fields up to approximately 1 nT.

The difference between the reconstructed current density $|\hat{\boldsymbol{J}}|$ and the corresponding FEM solution of $|\boldsymbol{J}|$ is more prominent. Although no noise was added to the simulated data, errors in the finite-difference approximations and artifacts in $\hat{\boldsymbol{B}}$ add up, so that the field estimate near the skull is highly distorted. The intra-cranial fields show greater resemblance, although a notable ringing-artifact from the skull can be seen in $|\hat{\boldsymbol{J}}|$.

Figure 9 displays the performance of the two ULF-MRI setups with the simulated imaging sequence described in section 4.1. **Figures 9A,B** show field reconstruction magnitude $|\hat{\boldsymbol{B}}_J|$ for a CDI sequence with 50 A polarizing current. The time-domain echo signals were superimposed with Gaussian noise of 2 fT/$\sqrt{\text{Hz}}$ and 0.35 fT/$\sqrt{\text{Hz}}$ for the multi-channel system and the single-channel system, respectively. The reconstruction quality is highly dependent on the SNR of the underlying ULF-MR images, which is shown in **Figures 9C,D**. With the ultra-sensitive single-channel setup, one achieves sensitivity in depth to the intra-cranial volume whereas the multi-channel setup gives a broader sensitivity pattern on the scalp and directly under the skull. **Figures 9E,F** illustrate estimates of the SNR maps of $\hat{\boldsymbol{B}}_J$, corresponding to the images in **Figures 9A,B**. The maps are derived from the noiseless $\hat{\boldsymbol{B}}_J$ and the SNR maps using Equation (19) with $c = 1.3$.

5. DISCUSSION

Hömmen et al. [9] concluded that an increase in image SNR of their setup is necessary for a successful *in-vivo* implementation of current-density imaging. However, based on measurements using simple phantoms, no exact numbers for the requirements in terms of SNR could be presented.

This work provides a profound understanding of the influence of noise on the reconstruction of the magnetic field $\boldsymbol{B}_J$ and the current density $\boldsymbol{J}$. The linearization of the field reconstruction gives an approximate relationship between the image SNR and the statistical uncertainty in the field estimates. Further, Monte-Carlo simulations were used to derive the statistical uncertainty in the presence of large background fields where the non-linearities take effect. The presented link between image SNR and noise in the reconstruction allows the determination of the necessary SNR for the reconstructions $\hat{\boldsymbol{B}}_J$ and $\hat{\boldsymbol{J}}$ within a predefined uncertainty. It also enables the assessment of the performance of specific ULF-MRI systems for zero-field-encoded CDI directly from acquired or simulated image data.

In order to retain constant image SNR in the Monte-Carlo simulations, we adjusted $|\boldsymbol{B}_\text{B}|$ to vary $\phi = \gamma\tau|\boldsymbol{B}_\text{B}|$. We set the zero-field-encoding time to $\tau = T_2$, which yields maximum SNR[$\boldsymbol{B}_J$] according to Vesanen et al. [6]. However, the non-linear dependence of SNR[$\boldsymbol{B}_J$] on ϕ suggests that there is an optimum set of parameters for each specific case. In reality, the effective background field will be roughly constant over the measurement periods and τ should be adjusted to obtain maximum SNR[$\boldsymbol{B}_J$]. If the relaxation times are known, Equations (19) and (20) can be utilized to create a cost function that provides parameters for maximum reconstruction quality. It should be mentioned that the optima for τ are flat and close to T_2 for small background fields. An adjustment of τ seems worthwhile in the case of very large background fields, where up to 12% can be gained in SNR[$\hat{\boldsymbol{B}}_J$] compared to $\tau = T_2$. Furthermore, it should be kept in mind that $\phi < \pi$ should be fulfilled to prevent ambiguity in the field reconstruction.

To analyze their performance and suitability for *in-vivo* CDI, our two ULF-MRI systems were examined in realistic image simulations. One was the system of Hömmen et al., including an optimized polarizing setup, and the second was a whole-head multi-channel system built at Aalto University. Key features that determine the SNR, such as the polarizing-field pattern, the coupling profile to the sensor, and noise, were accurately modeled. The estimates of $\boldsymbol{B}_J$ and $\boldsymbol{J}$ were derived from FEM simulations using a three-compartment head model. The peak current densities in intra-cranial tissue are similar to literature values, when scaled to the applied current of 4.5 mA [12, 13]. However, the three-compartment model neglects the fact that current is partly shunted by cerebrospinal fluid (CSF), which has a higher conductivity compared to gray- and white-matter tissue [13, 28].

The $\boldsymbol{B}_J$-field distribution served as an input for MRI simulations, emulating the entire sequence. Taking into account the insights from the Monte-Carlo simulations and the calculated SNR of the single-channel setup, the required improvement in SNR compared to Hömmen et al. [9] can now be specified. The simulations verify that the optimized polarization profile is sufficient. The peak SNR of the multi-channel setup is lower compared to the single-channel setup due to a higher sensor noise and different field coupling. A broader sample coverage, on the other hand, is provided by the multi-channel setup. The comparison between the two systems revealed that both high sensitivity and large sample coverage are required for current-density imaging usable for conductivity estimation.

It should be mentioned that both systems were evaluated with 50 A of polarizing current, which represents a close to maximum level for the room-temperature coil used with the single-channel device, whereas the superconducting polarizing coil used with the multi-channel device might be able to carry 2–4 times more current. Such an increase in the polarizing current benefits the image SNR and the SNR of the field estimates by the same factor. However, approaching such high fields will cause flux trapping in the sensor [18, 29, 30] and the superconducting filaments of the coil [17, 31], which has to be dealt with. Also larger currents required for the compensation of the field transient [32] can

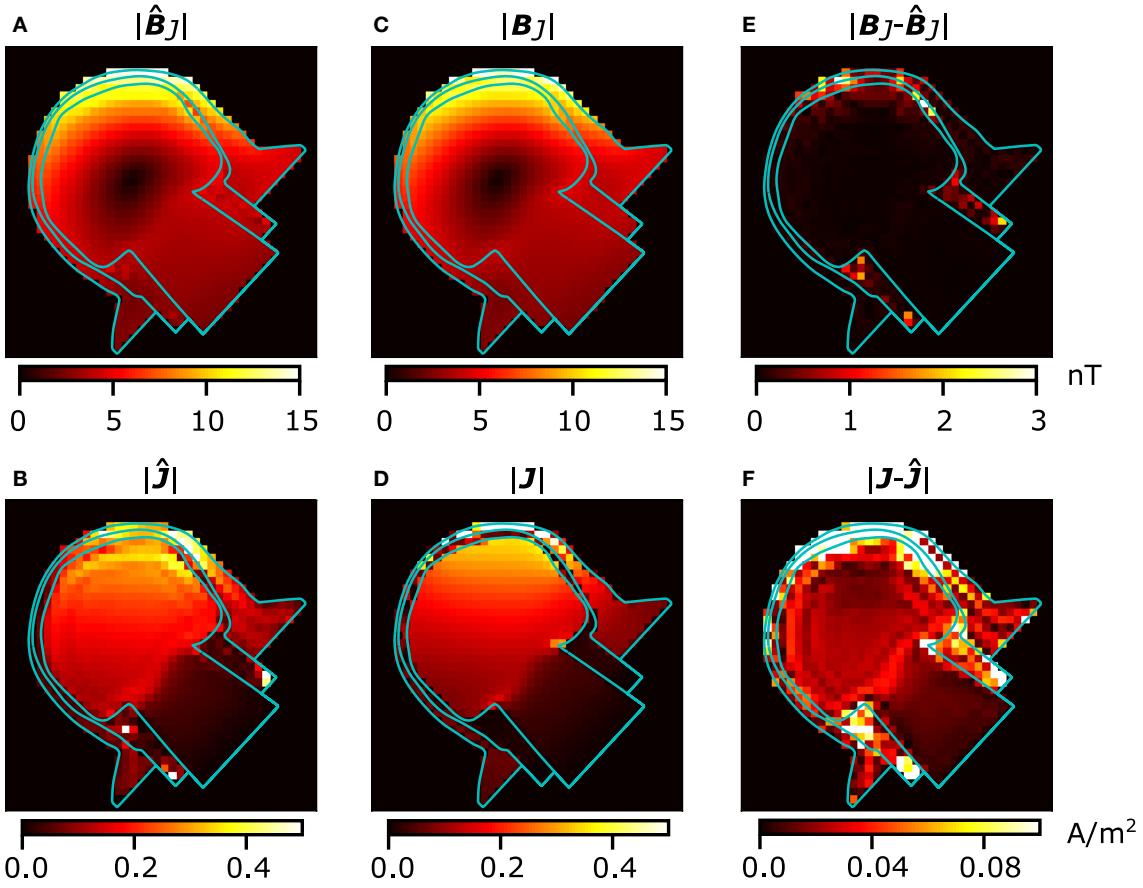


FIGURE 8 | Comparison of the simulated noiseless CDI reconstructions and the FEM solutions. The corresponding coordinates are given in **Figure 6B**. **(A)** shows the reconstructed magnetic field and **(B)** the reconstructed current density. **(C,D)** show the respective FEM solutions, and **(E,F)** display the absolute differences between reconstructions and the FEM solutions. The FEM fields are linearly interpolated from the FEM nodal values to the $(5 \times 5 \times 5)$ mm^3 voxel grid. The reconstructions are masked to zero outside the head model. Note that the color axes of the right-most figures differ from the others by a factor of 5.

cause excessive heating in the compensation coils, requiring more sophisticated techniques [33].

In addition, it should be mentioned that most tDCS devices do not support the application of 4.5-mA current. Nevertheless, some stimulators, such as DC-STIMULATOR PLUS (neuroConn, Germany), support 4.5-mA currents, provided that the electrode–skin resistance is low. The presented estimates for $\boldsymbol{J}$ and $\boldsymbol{B}_J$, as well as SNR[$\boldsymbol{B}_J$], scale linearly with the current strength. If the current was reduced to, e.g., 2 mA, the SNR of the single-channel system would still be sufficient to reconstruct $\boldsymbol{B}_J$ in the intra-cranial compartment. However, the reconstruction volume would be reduced. In case of the multi-channel system, the volume of reliable reconstruction would be limited mostly to the scalp compartment. With higher polarizing field, the reconstruction volume would, of course, be recovered again.

Besides noise, spatial leakage from the FFT has a significant influence on the quality of the reconstruction. Appropriate windowing of the *k*-space data manipulates the spatial response function of the voxels, effectively reducing the far-reaching leakage at the cost of a smoothed resolution. However, with the applied imaging and reconstruction procedures, leakage artifacts could not be entirely eliminated, yielding noticeable reconstruction errors, especially visible in the $\hat{\boldsymbol{J}}$-distribution. Besides spatial filtering, an effective method to reduce ringing artifacts in MRI is to apply more k-steps. However, this might not be applicable to *in-vivo* CDI as it would increase the measurement time significantly. Additionally, post-processing methods, for example "total variation constrained data extrapolation" [34], might reduce the artifacts without decreasing the image resolution.

The $\boldsymbol{J}$ reconstructions of both systems show limitations in thin tissue structures like the scalp. This is most probably due to the chosen resolution of $(5 \times 5 \times 5)$ mm^3, which does not allow sufficient gradient calculations in these areas. Reducing the voxel size to 1–2-mm would increase the quality of the $\hat{\boldsymbol{J}}$-distribution, but again at the cost of longer overall measurement time and lower SNR. Generally, the simulations show that the $\boldsymbol{B}_J$ reconstruction is more reliable than the $\boldsymbol{J}$ reconstruction, as artifacts strongly affect the gradient estimation.

Shall the reconstructions be used to fit individual conductivity values, superior results are expected when the $\hat{\boldsymbol{B}}_J$-field is used

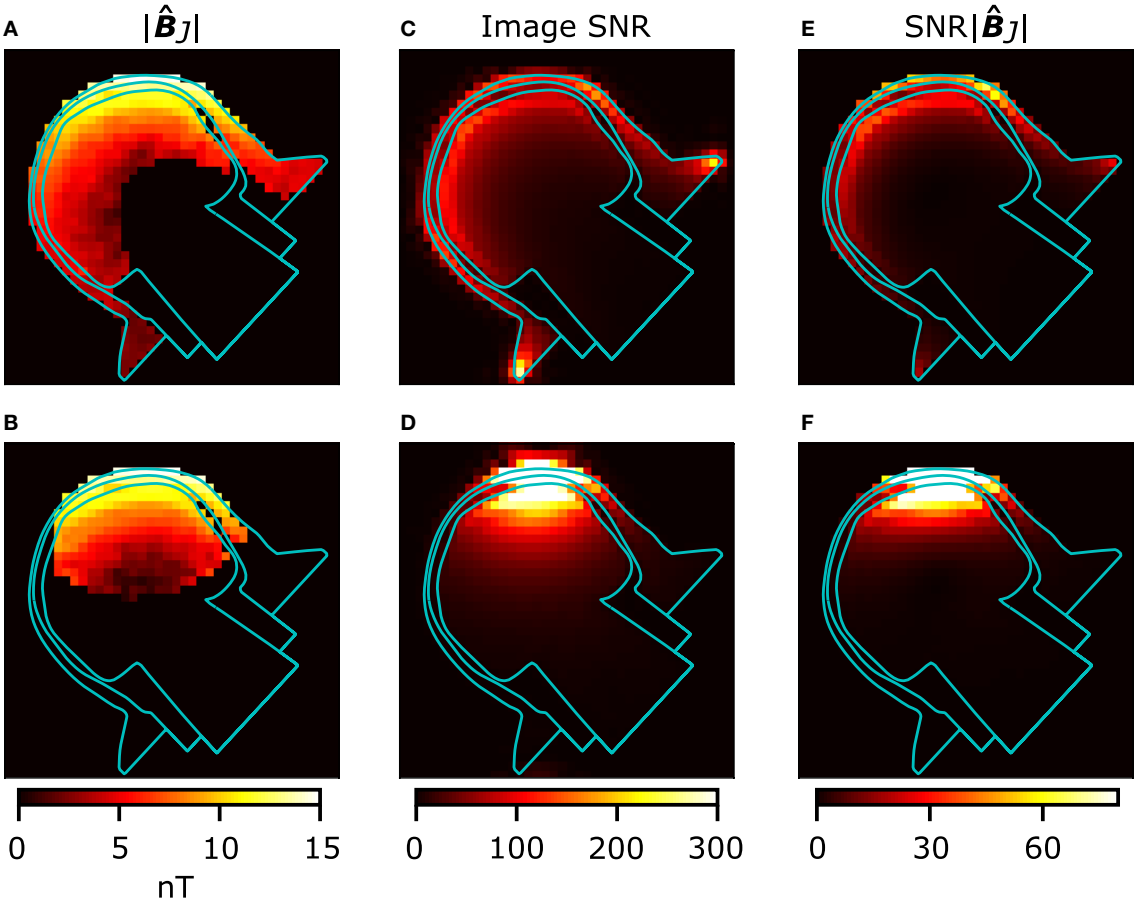


FIGURE 9 | Comparison of system performances of the Aalto multi-channel **(A,C,E)** and the PTB single-channel **(B,D,F)** ULF-MRI setups. **(A,B)** show $|\hat{B}_J|$ reconstructions of CDI simulations thresholded above SNR = 20 and inside the head model. **(C,D)** show the SNR maps of the simulated magnitude images and **(E,F)** contain estimates of SNRs of $|\hat{B}_J|$ in both systems.

as the measurement data. However, magnetic fields arising from the current leads should be either modeled or eliminated from the data. One way to exclude these fields would be to consider only closed path integrals of $\hat{B}_J$ and to apply the integral form of Ampère's law. It remains to be answered whether this enables to derive bulk conductivity values only, rather than spatially resolved conductivity mapping. Methods for this have not been presented so far and should be subject to further research.

6. CONCLUSION

We introduced methods to gain quantitative information about the effect of stochastic uncertainty on the non-linear reconstruction in zero-field-encoded current-density imaging (CDI). The work provides means to determine the ability of specific ultra-low-field MRI setups to reach acceptable signal-to-noise ratios in field reconstructions based on image SNR and to assess necessary improvements in, e.g., noise performance or polarizing field strength. By simulations, we evaluated the reconstruction quality of two existing setups under realistic conditions. We showed that current technology in ULF MRI is suitable for *in-vivo* CDI in terms of SNR. In addition, we encountered reconstruction errors due to a limited resolution and image artifacts requiring further research and development of more accurate reconstruction techniques.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

PH, AM, and RK contributed to the conception of the study. PH and AM performed the simulations, analyzed the results, wrote the first draft of the manuscript, and revised the manuscript based on the annotations of the co-authors. PH performed comparative measurements. AM contributed to the theory part in the **Appendix**. AH, RM, and JH developed the head model and performed FEM simulations. All the authors contributed

to manuscript revision and read and approved the submitted version. The study was supervised by JH, KZ, RI, and RK.

FUNDING

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 686865. It was partly supported by Vilho, Yrjö and Kalle Väisälä Foundation, by Project 2017 VF 0035 of the Free State of Thuringia, and by DFG Ha 2899/26-1.

ACKNOWLEDGMENTS

The authors thank Jan-Hendrik Storm for fruitful discussions on the FEM simulations.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2020.00105/full#supplementary-material

REFERENCES

1. Hämäläinen M, Hari R, Ilmoniemi RJ, Knuutila J, Lounasmaa OV. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev Modern Phys*. (1993) **65**:413–97. doi: 10.1103/RevModPhys.65.413
2. Vallaghé S, Clerc M. A global sensitivity analysis of three- and four-layer EEG conductivity models. *IEEE Trans Biomed Eng*. (2008) **56**:988–95. doi: 10.1109/TBME.2008.2009315
3. Opitz A, Windhoff M, Heidemann RM, Turner R, Thielscher A. How the brain tissue shapes the electric field induced by transcranial magnetic stimulation. *NeuroImage*. (2011) **58**:849–59. doi: 10.1016/j.neuroimage.2011.06.069
4. Nummenmaa A, Stenroos M, Ilmoniemi RJ, Okada YC, Hämäläinen MS, Raij T. Comparison of spherical and realistically shaped boundary element head models for transcranial magnetic stimulation navigation. *Clin Neurophysiol*. (2013) **124**:1995–2007. doi: 10.1016/j.clinph.2013.04.019
5. Miranda PC, Callejón-Leblic MA, Salvador R, Ruffini G. Realistic modeling of transcranial current stimulation: the electric field in the brain. *Curr Opin Biomed Eng*. (2018) **8**:20–7. doi: 10.1016/j.cobme.2018.09.002
6. Vesanen PT, Nieminen JO, Zevenhoven KCJ, Hsu YC, Ilmoniemi RJ. Current-density imaging using ultra-low-field MRI with zero-field encoding. *Magn Reson Imaging*. (2014) **32**:766–70. doi: 10.1016/j.mri.2014.01.012
7. Nieminen JO, Zevenhoven KCJ, Vesanen PT, Hsu YC, Ilmoniemi RJ. Current-density imaging using ultra-low-field MRI with adiabatic pulses. *Magn Reson Imaging*. (2014) **32**:54–9. doi: 10.1016/j.mri.2013.07.012
8. Zevenhoven KCJ, Alanko S. Ultra-low-noise amplifier for ultra-low-field MRI main field and gradients. *J. Phys.: Conf. Ser*. (2014) **507**:042050. doi: 10.1088/1742-6596/507/4/042050
9. Hömmen P, Storm JH, Höfner N, Körber R. Demonstration of full tensor current density imaging using ultra-low field MRI. *Magn Reson Imaging*. (2019) **60**:137–44. doi: 10.1016/j.mri.2019.03.010
10. Bikson M, Grossman P, Thomas C, Zannou AL, Jiang J, Adnan T, et al. Safety of transcranial direct current stimulation: evidence based update 2016. *Brain Stimul*. (2016) **9**:641–61. doi: 10.1016/j.brs.2016.06.004
11. Antal A, Alekseichuk I, Bikson M, Brockmöller J, Brunoni AR, Chen R, et al. Low intensity transcranial electric stimulation: safety, ethical, legal regulatory and application guidelines. *Clin Neurophysiol*. (2017) **128**:1774–809. doi: 10.1016/j.clinph.2017.06.001
12. Miranda PC, Lomarev M, Hallett M. Modeling the current distribution during transcranial direct current stimulation. *Clin Neurophysiol*. (2006) **117**:1623–9. doi: 10.1016/j.clinph.2006.04.009
13. Neuling T, Wagner S, Wolters CH, Zaehle T, Herrmann CS. Finite-element model predicts current density distribution for clinical applications of tDCS and tACS. *Front Psychiatry*. (2012) **3**:83. doi: 10.3389/fpsyt.2012.00083
14. Espy MA, Magnelind PE, Matlashov AN, Newman SG, Sandin HJ, Schultz LJ, et al. Progress toward a deployable SQUID-based ultra-low field MRI system for anatomical imaging. *IEEE Trans Appl Superconduct*. (2015) **25**:1–5. doi: 10.1109/TASC.2014.2365473
15. Inglis B, Buckenmaier K, SanGiorgio P, Pedersen AF, Nichols MA, Clarke J. MRI of the human brain at 130 microtesla. *Proc Natl Acad Sci USA*. (2013) **110**:19194–201. doi: 10.1073/pnas.1319334110
16. Vesanen PT, Nieminen JO, Zevenhoven KCJ, Dabek J, Parkkonen LT, Zhdanov AV, et al. Hybrid ultra-low-field MRI and magnetoencephalography system based on a commercial whole-head neuromagnetometer. *Magn Reson Med*. (2013) **69**:1795–804. doi: 10.1002/mrm.24413
17. Lehto I. *Superconducting Prepolarization Coil for Ultra-Low-Field MRI*. Aalto University School of Science (2017). Available online at: http://urn.fi/URN:NBN:fi:aalto-201712188119
18. Luomahaara J, Kiviranta M, Grönberg L, Zevenhoven KC, Laine P. Unshielded SQUID sensors for ultra-low-field magnetic resonance imaging. *IEEE Trans Appl Superconduct*. (2018) **28**:1–4. doi: 10.1109/TASC.2018.2791022
19. Kraus R Jr, Espy M, Magnelind P, Volegov P. *Ultra-Low Field Nuclear Magnetic Resonance: A New MRI Regime*. New York, NY: Oxford University Press (2014). doi: 10.1093/med/9780199796434.001.0001
20. Zevenhoven KCJ, Mäkinen AJ, Ilmoniemi RJ. Superconducting receiver arrays for magnetic resonance imaging. *Biomed Phys Eng Exp*. (2020) **6**:015016. doi: 10.1088/2057-1976/ab5c61
21. Brown RW, Cheng YCN, Haacke EM, Thompson MR, Venkatesan R. *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. Hoboken, NJ: John Wiley & Sons (2014). doi: 10.1002/9781118633953
22. Pruessmann KP. Encoding and reconstruction in parallel MRI. *NMR Biomed*. (2006) **19**:288–99. doi: 10.1002/nbm.1042
23. Löwdin PO. On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals. *J Chem Phys*. (1950) **18**:365–75. doi: 10.1063/1.1747632
24. Aiken JG, Erdos JA, Goldstein JA. On Löwdin orthogonalization. *Int J Quant Chem*. (1980) **18**:1101–8. doi: 10.1002/qua.560180416
25. Hanson JD, Hirshman SP. Compact expressions for the Biot–Savart fields of a filamentary segment. *Phys Plasmas*. (2002) **9**:4410–2. doi: 10.1063/1.1507589
26. Hunold A, Güllmar D, Haueisen J. *CT Dataset of a Human Head*. Zenodo (2019). Available online at: https://zenodo.org/record/3374839#.XXZNtHtCSUk
27. Zotev VS, Matlashov AN, Savukov IM, Owens T, Volegov PL, Gomez JJ, et al. SQUID-based microtesla MRI for in vivo relaxometry of the human brain. *IEEE Trans Appl Superconduct*. (2009) **19**:823–6. doi: 10.1109/TASC.2009.2018764
28. Salvador R, Mekonnen A, Ruffini G, Miranda PC. Modeling the electric field induced in a high resolution realistic head model during transcranial current stimulation. In: *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. (Buenos Aires) (2010). p. 2073–6. doi: 10.1109/IEMBS.2010.5626315
29. Luomahaara J, Vesanen P, Penttilä J, Nieminen J, Dabek J, Simola J, et al. All-planar SQUIDs and pickup coils for combined MEG and MRI. *Superconduct Sci Technol*. (2011) **24**:075020. doi: 10.1088/0953-2048/24/7/075020
30. Al-Dabbagh E, Storm JH, Körber R. Ultra-sensitive SQUID systems for pulsed fields—degaussing superconducting pick-up coils. *IEEE Trans Appl Superconduct*. (2018) **28**:1–5. doi: 10.1109/TASC.2018.2797544
31. Zevenhoven KCJ. *Solving Transient Problems in Ultra-Low-Field MRI*. University of California; Berkeley and Aalto University (2011). Available online at: http://urn.fi/URN:NBN:fi:aalto-201305163099

32. Nieminen JO, Vesanen PT, Zevenhoven KCJ, Dabek J, Hassel J, Luomahaara J, et al. Avoiding eddy-current problems in ultra-low-field MRI with self-shielded polarizing coils. *J Magn Reson.* (2011) **212**:154–60. doi: 10.1016/j.jmr.2011.06.022
33. Zevenhoven KCJ, Dong H, Ilmoniemi RJ, Clarke J. Dynamical cancellation of pulse-induced transients in a metallic shielded room for ultra-low-field magnetic resonance imaging. *Appl Phys Lett.* (2015) **106**:034101. doi: 10.1063/1.4906058
34. Block KT, Uecker M, Frahm J. Suppression of MRI truncation artifacts using total variation constrained data extrapolation. *Int J Biomed Imaging.* (2008) **2008**:1–8. doi: 10.1155/2008/184123

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 15 April 2020
doi: 10.3389/fphy.2020.00080

Anatomically Adaptive Coils for MRI—A 6-Channel Array for Knee Imaging at 1.5 Tesla

Bernhard Gruber[1,2]*, *Robert Rehner*[3], *Elmar Laistler*[1] *and Stephan Zink*[3]

[1] Division MR Physics, Center for Medical Physics and Biomedical Engineering, Medical University Vienna, Vienna, Austria, [2] A.A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA, United States, [3] Siemens Healthcare GmbH, Erlangen, Germany

Purpose: Many of today's MR coils are still somehow rigid and inflexible in their size and shape as they are intentionally designed to image a specific anatomical region and to fit a wide range of patients. Adaptive coils on the other hand, are intended to follow a one-size-fits-all approach, by fitting different shapes, and sizes. Such coils improve the SNR for a wide range of subjects by an optimal fit to the anatomical region of interest, and in addition allow an increased handling and patient comfort as one MRI receive-coil is maintained instead of multiple.

Material and Methods: To overcome the SNR losses by non-fitting and thus poorly loaded RF coils, we propose a stretchable antenna design. Each loop has the ability to reversibly stretch up to 100% of its original size, to be anatomically adaptive to different shapes and sizes, and therefore make the coil usable for a wide patient population. Besides the mechanical challenge to find a robust but flexible conductive material, various other problems like frequency and matching shifts affect the SNR. Through bench measurements and MR Imaging at 1.5 T, we investigated different stretchable conductor materials, that fit the defined requirements. Finally, a rigid reference coil and an adaptive 6-channel array for knee imaging at 1.5 Tesla were developed to investigate the potential improvement in SNR.

Results: The material tests identified two potentially useful materials: Highly ductile copper and a silver-plated stranded copper wire. Although, the adaptivity causes a frequency shift of the resonance frequency, which entails in variations of the impedance that each coil presents to its connected pre-amplifier, there are strategies to mitigate these effects. The adaptive array prototype made of partly-stretchable loops, showed an improved SNR of up to 100% in 20 mm depth from the phantom surface, and therefore demonstrates the effectiveness of adaptive coils.

Keywords: adaptive coils, stretchable loop, meandered conductor, SNR, one-size-fits all

OPEN ACCESS

Edited by:
Simo Saarakkala,
University of Oulu, Finland

Reviewed by:
Ji Chen,
University of Houston, United States
Manuel José Freire Rosales,
University of Seville, Spain

***Correspondence:**
Bernhard Gruber
b.gruber@ieee.org

Specialty section:
This article was submitted to Medical Physics and Imaging, a section of the journal Frontiers in Physics

Received: *29 November 2019*
Accepted: *09 March 2020*
Published: *15 April 2020*

Citation:
Gruber B, Rehner R, Laistler E and Zink S (2020) Anatomically Adaptive Coils for MRI—A 6-Channel Array for Knee Imaging at 1.5 Tesla. Front. Phys. 8:80. doi: 10.3389/fphy.2020.00080

INTRODUCTION

The past two decades of Magnetic Resonance Imaging (MRI) have seen immense advances in various fields, with a focus toward improved sensitivity, multi-modal imaging and of course reduced scan-time in clinical and research examinations. Acquiring MRI data is still time consuming due to long acquisition times, and therefore prone to motion artifacts. Furthermore,

MRI data acquisition is limited in spatial and temporal resolution due to the lack of signal-to-noise ratio (SNR). A simple solution is to apply higher static magnetic field strength (B_0) [1] to increase the detectable nuclear magnetization, and thus to achieve higher spatial resolution with sufficient high SNR [2].

While the advancement of gradient coils in strength and slew-rate [1] ensured a speed up in image acquisition, the improvement of sensitivity with higher field strengths or well-crafted detector geometries of MRI probes, have always been critical [3].

Back in 1980, Ackerman et al. demonstrated that an improved SNR could be obtained by placing a small coil on the surface of the sample, close to the region of interest [4]. The use of small surface coils in the regime of sample dominated noise enables large sensitivity improvements, because it provides both, stronger magnetic coupling with the sample and noise reduction due to the smaller volume of tissue being visible for the coil [5].

Many theoretical and experimental works suggested to put a large number of small surface coils as close as possible to the imaging volume to achieve a set of advantageous features like a high filling factor [6–9]. Because of the importance of coil detectors, and the fact that MRI (today) relies on signal detection with receive arrays, the development of such RF antennas is a critical step in gaining SNR, speeding up the acquisition and therefore improving the patient comfort during every MR examination [10–13]. To achieve a high SNR, it is important that at high Q_{loaded} values, the filling factor is very close to $\eta = 1$, to gain the maximum SNR. The magnetic field filling factor η is a ratio, for the magnetic energy stored inside a load (e.g., sample, phantom, patient) to the total magnetic energy stored in the coil.

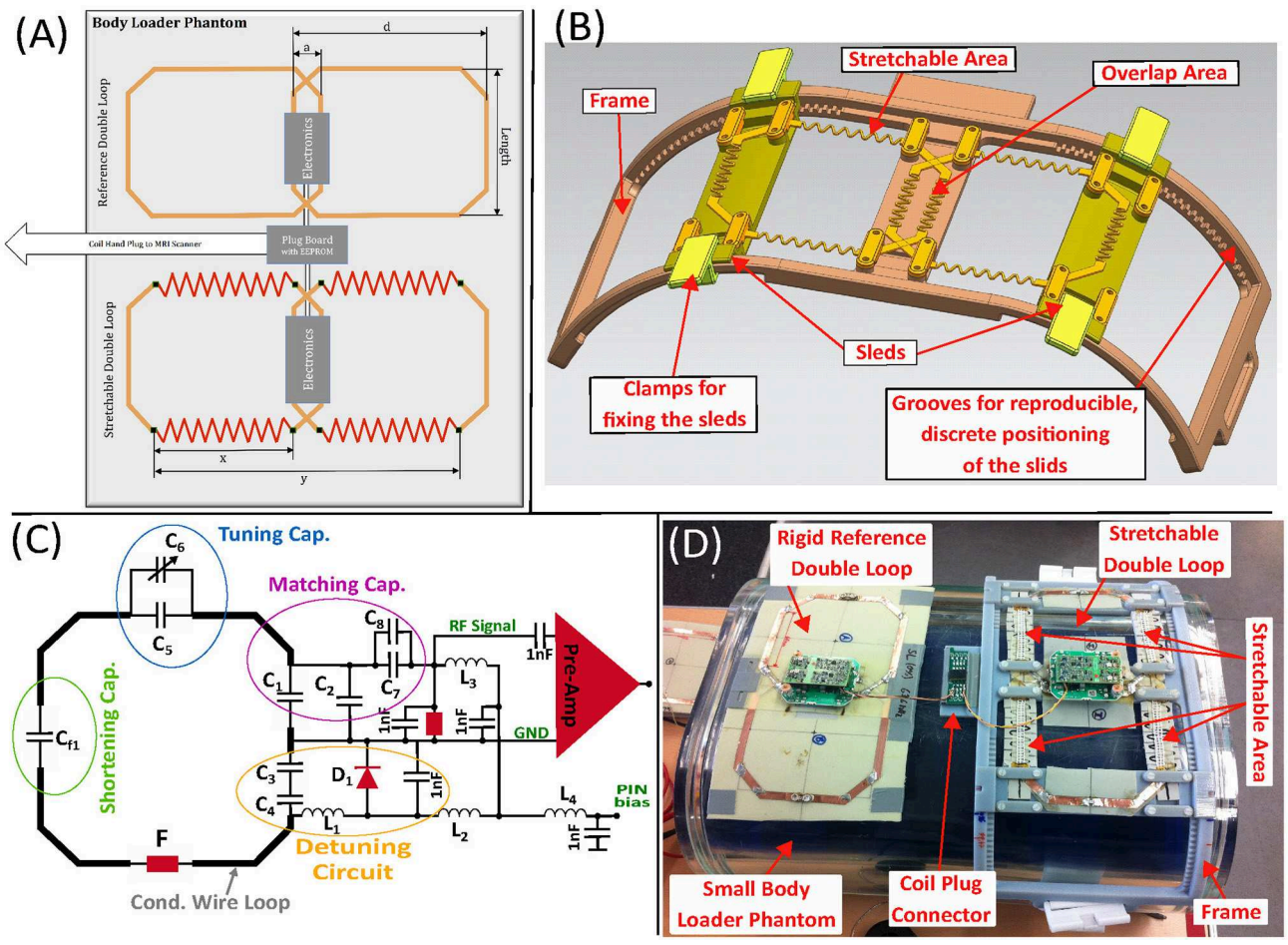


FIGURE 1 | (A) Schematic of the used test-setup for material measurements. The setup allows reproducible testing of different conductive materials used for an adaptive double-loop configuration that can be enlarged/stretched and immediately compared to a rigid reference double-loop. **(B)** Two rectangular loops with a maximum length (z-direction) of 100 mm can be attached to the frame (brown) and enlarged in x-direction by shifting the sleds (yellow) in 5 mm steps. **(C)** Simplified RF circuit. **(D)** Experimental setup with a rigid reference double-loop and a stretchable double-loop mounted on the test-frame, placed on the torso-shaped phantom (see Figure 6) with a main body (blue) filled with per 1,000 mL Bayol-oil and 0.011 g MACROLEX blue and an outer compartment (clear) filled with per 1,000 g H_2O dist.: 1.25 g $NiSO_4x6H_2O$, 5 g NaCl. The coil plug connector in between the two double-loops serves as connector between the test setup and the MRI scanner and houses the coil code. The test-frame is 3D laser-sintered (*StratasysObjetEden500V*) using MR-invisible material.

$$\eta_f = \frac{\int_{Sample\ Volume} B_1^2 dV}{\int_{Total\ Volume} B_1^2 dV} \quad (1)$$

B_1 is the value of the RF magnetic field once integrated over the sample and the second time integrated over the total coil volume. As an approximation η_f can also be seen as:

$$\eta_f \approx \frac{B_1^2}{Q_R P} \quad (2)$$

where

$$Q_R = \frac{Q_{unloaded}}{Q_{loaded}} \quad (3)$$

is the unloaded Q-value divided by the loaded Q-value of the coil element and P is the input RF power. The introduction of a load decreases the quality factor of the coil and the magnetic field [14]. By bringing the coil closer to the sample, the filling factor is increased and the term $\sqrt{2}B_t$ in the numerator of the equation for the SNR,

$$SNR = \frac{U_{Signal}}{U_{Noise}} = \frac{\sqrt{2}\omega \Delta V M_{xy} |B_t|}{\sqrt{4kT_{eff}\Delta f R_{eff}}} \quad (4)$$

is maximized, which results in an optimized SNR. Many factors determine the SNR available in an MR experiment. In Equation 4 the U_{Signal} considers the SNR for a single voxel volume ΔV, with the assumption that the fields of the magnet and the coils are constant over the voxel. The properties of the sample and the coil, contribute to the SNR through the resistance at the coil terminals (R_{eff}) and the sensitivity pattern of the coil. The noise signal U_{Noise} in any MRI experiment is basically thermal noise generated

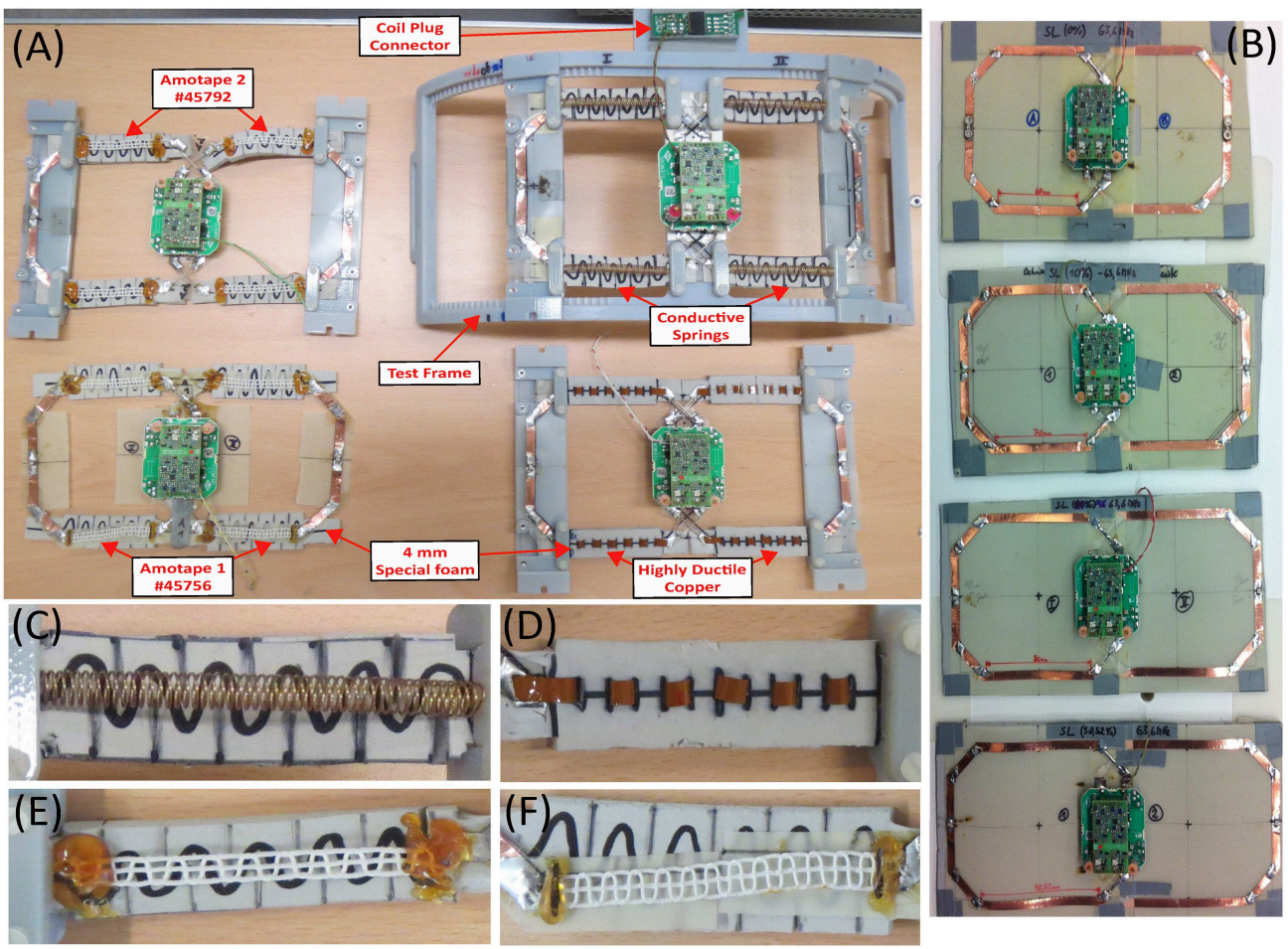


FIGURE 2 | (A) The four investigated stretchable material candidates. A highly reversible flexible special foam padding was used, to support the stretchable materials. The pre-amplifier on the feed boards are already attached. **(B)** Four differently sized rigid double-loops used as a reference. **(C)** Stretchable double-loop made of CuBe-Strain springs. **(D)** Highly ductile copper AP9121R DuPont™Pyralux® flexible laminate. The material is a double-sided, 35 µm copper-clad flexible laminate. **(E)** Amotape® Conduct Elast. *#45756*. A 6 mm wide stranded copper wire made of 7 single strands with an area of 0.078 mm^2 each. PTFE insulated and connected with 3 elastic threads made of Elastane gimped with PA66 (620dtex + 78/2dtex). **(F)** Amotape® Conduct Elast. *#45792* with same specifications as Amotape® #45756 but with 19 single strands.

by the receiver coil and the sample scaled to the bandwidth used in detecting the signal [15].

So, by improving the filling factor, the coil resistance R_{Coil}, also called *equivalent-series-resistance (ESR)*, is minimized by making the unloaded Q high, and the sample resistance R_s (through the induced eddy current losses in the conductive sample) can be minimized by choosing the coil size to match the target Field of View (FoV).

Designing a coil array to fit close to the region of interest is quite easily achieved, but to use the same coil with multiple patients that vary in size and shape, is challenging. It requires the coil array to be shape/form adaptive. Mechanical flexibility of the RF array is advantageous as in most cases the shape and size of different body anatomies varies significantly, but most RF coils are rigid and only fit a specific anatomical region and only certain patient sizes. Form-adaptive RF coil arrays improve the electromagnetic coupling between sample and coil, provide a higher filling factor, and therefore, potentially improve the RF receive efficiency, if the mismatch and expected frequency shift is not exceeding the benefits.

Allowing adjustable coil geometries requires equally flexible solutions for mitigating upcoming parasitic effect like increased mutual coupling between elements and frequency shifts. As a result of these shifts in center frequency and the variation in coupling, the source impedance presented to the pre-amplifier changes, which leads to SNR loss. Several approaches to mitigate the effects of coil coupling and frequency shifts, like broad-/wide -band matching [16] have already been more or less implemented in the field. Another approach is to automatically tune and match the coil array [17–20]. Previous work have already shown the feasibility of mechanical adaptation of the receive coil to the body part of interest. A first approach of flexible coils, where a mercury filled tube by Malko et al. [21] was used to form a loop. Other works on adaptive coils include a transmit-receive head array that permits bending to adjust its diameter [22], a sliding mechanism varying the diameter of a conical coil arrangement for wrist imaging [23]. A stretchable coil array for knee imaging at 3T, which utilized braided copper wire mounted on an elastic textile substrate was introduced cite Nordmeyer-Massner et al. [23]. Compared to a standard rigid knee array, the stretchable 8 channel array introduced an overall SNR loss of 20%. Later on on-coil digitization avoiding cabling and increase patient comfort [24] was used with the stretchable copper braid. Further approaches focused on mechanical flexibility, which is offered by several coil designs like screen-printed flexible MRI receive coils [25], or the flexible/rigid PCBs [26, 27]. Ongoing research on coil elements made of coaxial cable looks promising, especially due to the low inter-element coupling [28–30]. Such elements offer a large range in flexibility, potentially enabling wearable coil arrays, but limitations like the dependency of the resonance frequency to the permittivity of the dielectric, and therefore to the diameter of coaxial cable introduce limitations that need to be further investigated.

The present work addresses mechanical and electrical issues of anatomically adaptive coil arrays. In the first part, the concept of partly-stretchable coil arrays is investigated alongside with material tests of conductive materials used as coil elements. In the second part, the performance of a 6-channel adaptive array for 1.5 T is compared with a rigid reference array (Similar to the 15-channel Tx/Rx knee coil array from QED). Both arrays were tested using three different sized knee phantoms, representing various realistic knee sizes. Parts of this work have already been presented at conferences [31, 32].

MATERIALS AND METHODS

Material Tests

The key challenge in designing anatomically adaptive loops is to find reversibly stretchable coil conductors, which have equal or only slightly worse properties than the used standard (flat) wire with respect to electrical conductivity.

For typical metallic wire, stretching results in irreversible plastic deformation beyond a few percent of elongation. A certain reversibility of a change in length can be achieved by mounting a wire with a certain length reserve (e.g., meandered style) on an elastic material. In addition, this material should

TABLE 1 | Measurement points and fixed/calculated geometrical values for the experimental setup.

Stretching [%]	x [mm]	y [mm]	d [mm]	a [mm]	Length z-direction [mm]
0	60	190	100	14.0	100
10	70	209	110	15.4	100
20	80	228	120	16.8	100
30	90	247	130	18.2	100

*x is the length of the stretchable segment. y is the distance between the ends of the two stretchable segments of the double loop. d is the total width of one element. a is the overlap distance between the two coils (see also **Figure 1A**).*

TABLE 2 | Materials used for stretchable areas. Amotape® #45756 and Amotape® #45792 provide the stretchability of the stranded wire by three elastic Elastane threads.

Designation	Properties
Amotape®Conduct Elast. *#45756*	6 mm wide stranded copper wire consisting of 7 single strands with an area of 0.078 mm^2 each; PTFE insulation around wire 3 elastic threads made of Elastane gimped with PA66
Amotape®Conduct Elast. *#45792*	as above, but 19 single strands
Highly Ductile Copper—*AP9121R* DuPont™ Pyralux© AP flexible laminate	Doubled-sided, copper-clad laminate; polyamide composite copper foil with 0.0508 mm dielectric thickness and 35 μm copper thickness
$CuBe_2$ strain-spring	Wire thickness $d = 0.5\ mm$; Outer diameter $D_e = 5.9\ mm$; Inner diameter $D_i = 4.9\ mm$; Length $L_0 = 62\ mm$; Windings $n = 50$, overall wire length $L_a = 882.16\ mm$; tensile strength 950 N/mm^2

*The wire is woven into the threads in a meandered way (see **Figures 2E,F**) and allows a 50% reversible elongation. The highly ductile copper (see **Figure 2D**) is not stretchable per se, but provides high flexibility without any drawbacks in conductivity through its wave-like arrangement within a foam sheet of 5 mm thickness. The last material tested are $CuBe_2$ strain-springs (see **Figure 2C**).*

be "MR silent," which means that there is no contribution to the MR signal from this material. To restore the original state after stretching, the wire has to allow enough elasticity and ideally should mechanically behave like a spring. Standard highly conductive materials like copper, silver, gold, aluminum were too pliable for this task. Materials like iron or steel would fulfill the strength requirement, but they are ferromagnetic, and therefore not suitable for MRI. An option in between would be austenitic steel, which is not ferromagnetic and would provide enough strength, but has low conductivity. We developed several mechanical concepts, published in a filed patent [33]. One of the most promising approaches thereof is the *partly-stretchable-loop*—concept (see **Figures 2A,C–F**) used in this work.

Four different reversibly stretchable materials were investigated (see **Table 2**) and used for the construction of stretchable double-loops (see **Figure 2A**). The loops were realized with a stretchable area of length x and a width of 100 mm (z-direction) (see **Figure 1A**). The stretchable double-loops with loop sizes according to **Table 1** were compared to four standard double-loops (see **Figure 2B**) as reference.

The double-loop arrays with rectangular loops were manufactured on FR-4 (fiberglass cloth with a flame-resistant epoxy resin) using the simplified circuitry illustrated in **Figure 1C**.

The reference loops and the fixed parts of the partly-stretchable loops were made of 6 mm wide adhesive copper tape with a copper thickness of 70 μm. Thin or very narrow copper traces increase the loops resistance thus lowering the Q value, while a wider or thicker copper trace may cause eddy current heating, B_1-distortions and/or self-shielding [14]. 16-awg thick tin-plated copper wire bridges were used to overlap at the cross sections of each loop, minimizing capacitance between the two loop traces. Segmenting capacitors were used to provide a homogeneous current distribution over the loop, reduce the E-fields induced into the sample load through voltage splitting between the capacitors, reduce the stray fields caused by the split voltages which influence the load dependence of the resonance frequency and finally reduce capacitive coupling as well as parasitic capacitance between loop and sample. Each loop was tuned to the resonance frequency of 63.6 MHz (1.5T) using a torso-shaped phantom with a main body (blue: per

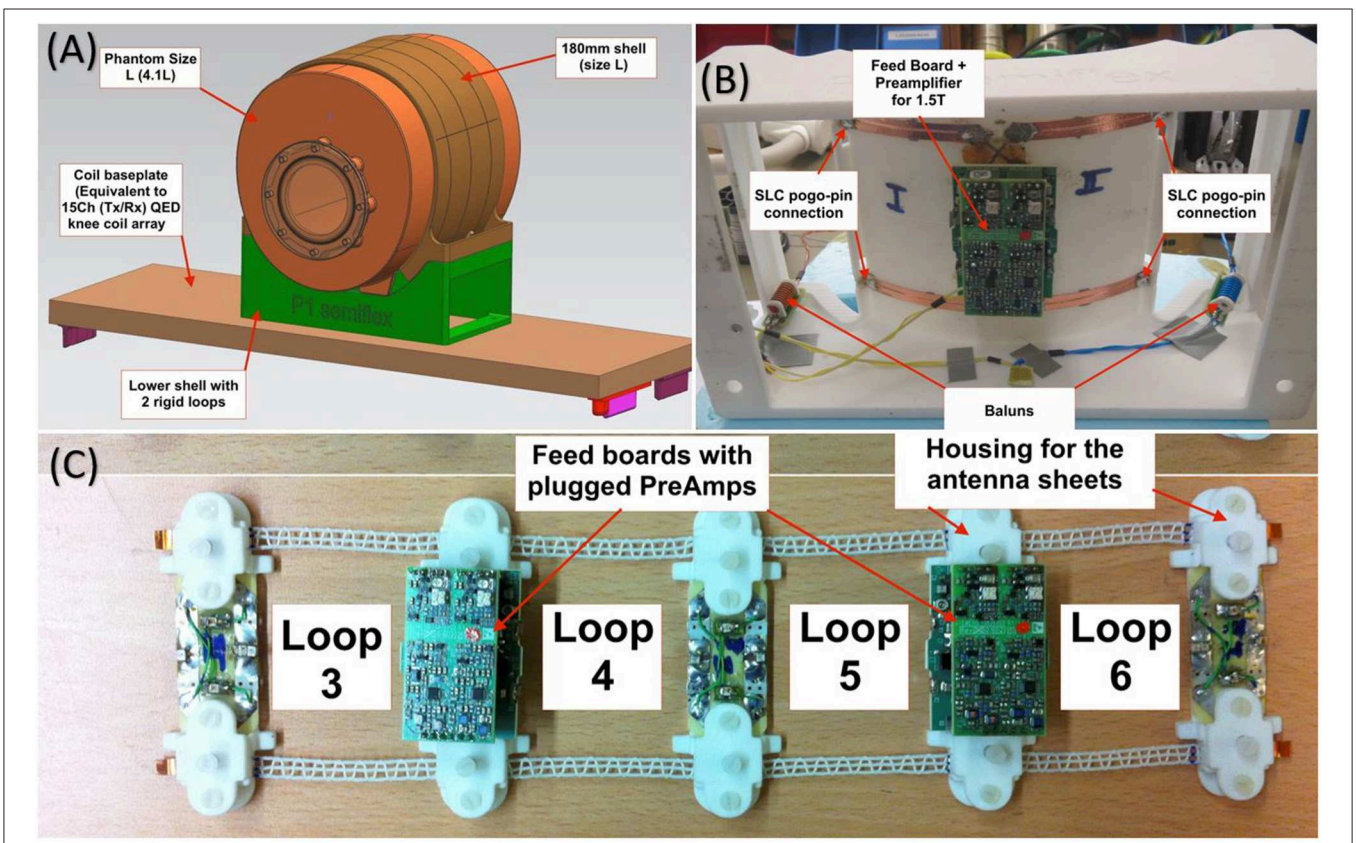


FIGURE 3 | (A) 3D rendering of the reference array mounted on the lower shell with the largest knee Phantom L (4.1 l). **(B)** The lower shell houses two channels that can be electrically connected to the upper four channels using spring-loaded pins melted into the laser-sintered lower shell. Such *pogo-pins* (preci-dip SA, Delemont, Switzerland) offer a low resistance military standard connection with high mating cycles and excellent performance at higher frequencies like MRI. The lower shell contains cable traps (blue and yellow) to reduce common mode currents on the shield. **(C)** 4-channel adaptive array with 4 stretchable elements, were each loop is made of Amotape® #45792, with a stretchable area between the feed-boards of x = 60.1 mm. The 4-channel adaptive array is connected to the 2-channel lower shell for measurements. Together they make the 6-channel adaptive array. The lower shell can be used along with the 4-channel adaptive array or the 4-channel rigid array.

1,000 mL Bayol-oil and 0.011 g MACROLEX blue) and an outer compartment (clear: per 1,000 g H_2O dist.: 1.25 g $NiSO_4 \times 6H_2O$, 5g NaCl) (see **Figure 1D**).

The stretchable double-loops were mounted on a 3D laser-sintered (ObjetEden500V, Stratasys, Rechovot, Israel) frame (see **Figures 1B,D**), which allows the stretching of the loop material in discrete steps of 5 mm in x-direction (the stretching in y-direction is also performed, as the frame is attached to the phantom's surface). The overlap between two loops is constant and provides optimal overlap decoupling for a stretching of 15% (x = 75 mm).

Bench and MRI measurements were performed with 0, 10, 20, and 30% stretching (see **Table 1**). The stretch area x is fixed to start at 60 mm. The geometrical properties are calculated with Equation 5 and 6, which result from an array design of a 18-channel adaptive knee array with 3 rows of 6 rectangular loop elements (see **Table 1** line 1).

$$\% - Stretching = \frac{y_n - y_0}{\frac{y_0}{100}} \tag{5}$$

$$x = \frac{y}{1.9} - 40\ mm \tag{6}$$

Stretching of the adaptive loops changes the loop's inductance and resistance and is therefore expected to cause a shift in the resonance frequency. Therefore, the partly-stretchable loops were tuned and matched to the Larmor frequency at 1.5 Tesla (63.6 MHz) at the center between the maximum and minimum length of the stretchable areas corresponding to a stretching of 15% with $x = 75$ mm). The four reference double-loops were individually tuned to and matched. Capacitor values for all built double-loops can be seen in **Figure 12**.

The inductance of each loop is estimated from the total capacitance and the resonance frequency. The coil resistance is estimated from the measured Q-factor. Q-values for each loop were measured at a distance of 20 mm with an S21 measurement on the network analyzer (E5071C, Keysight Technologies, Santa Rosa, CA, USA) using a double-loop probe with −75 dB decoupling. To estimate the effect of stretching on inter-element coupling, the coupling coefficient k was measured using the two-mode-frequencies method:

$$k = \frac{f_{in-phase}^2 + f_{anti-phase}^2}{f_{in-phase}^2 - f_{anti-phase}^2} \tag{7}$$

with $f_{in-phase} = \frac{1}{2\pi\sqrt{C^*(L-M)}}$ and $f_{anti-phase} = \frac{1}{2\pi\sqrt{C^*(L+M)}}$ [34, 35]. The upper frequency and the lower frequency mode are measured on the network analyzer, with an S21 measurement, assuming that the resonant frequencies f_0 are identical for both loops, as well as the capacitance and the inductance values for both loops are the same. The center-to-center distance for rectangular loops to achieve optimal inductive decoupling is 0.9 d [6], but due to the stretching of elements, the decoupling between elements changes. Residual coupling was suppressed by pre-amplifier decoupling. Active detuning using an LC parallel circuit was implemented to detune the loop during transmit, and baluns, reducing common mode currents on the shield of the coaxial cables, were added and tuned to the Larmor frequency.

The experimental setup (see **Figure 1D**) was placed in the MRI and with every stretchable double-loop configuration and the corresponding reference double-loop noise and signal data was acquired one after another. For every configuration, 2 measurements were performed: per stretched position (4 stretch-points, see **Table 1**, column 1) signal and noise data was acquired with the reference and the stretchable double loops.

SNR images were generated from signal- and noise datasets, acquired on a Siemens MAGNETOM Aera 1.5 Tesla MRI scanner with software platform Syngo MR E11 (Siemens Healthcare GmbH, Erlangen, Germany). To obtain the SNR images, a standard spin-echo sequence (TE = 15 ms, TR = 300 ms, FoV = 300 mm, TA = 1:20 min, slice thickness = 5 mm, acq. matrix = 256 × 256, voxel = 1.2 × 1.2 mm, bandwidth = 130 Hz/pixel) with a 90° and a 180° RF pulse was applied. For the noise measurement, the RF excitation pulse was set to zero, whereas for the signal measurement the RF excitation was set automatically. The acquired data were exported and then reconstructed offline MATLAB (MATLAB, The Mathworks, Natick, MA, USA).

SNR was calculated using the ***Sum-of-Squares (SoS)*** method ideal for high input SNR [36] and the ***Maximum Available (MA)*** method [15].

$$SNR_{SoS} = 2 * \frac{\rho^2}{\sigma^2} * \sum_{k=1}^{N} |c_k|^2 \tag{8}$$

SNR_{SoS} values as calculated in Equation 8, are equal to the SNR for optimal combining with unknown coil sensitivities. The MA method describes optimal coil combination, were sensitivities are known. By multiplying the pixel value in each coil with the complex conjugate of the coil sensitivity for that channel, summing over all channels, and dividing this sum with the sum of the squared coil sensitivity in all channels, an optimal noise decorrelated (noise pre-whitened) combination method is used. Any phase added by the coil itself is removed and the signal is summed up. This method is also known as B_1-weighted coil combination [6].

Adaptive Knee Array

Based on the results of the material test, the best performing material was used for further investigation. A preliminary study of the knee geometry (*100 mm up/down the knee center*) on 25 patients from Europe and U.S.A. showed a diameter range of 102–169 mm for European knees and 110–211 mm for U.S. knees. Standard coils like the 15-channel Tx/Rx knee coil by QED (Quality Electrodynamics, Mayfield village, Ohio, USA) have a limited inner diameter of 173 mm and field of View (FoV) in z-direction of around 200 mm. This diameter limitation makes the knee a good object to demonstrate the performance of a ***one-size-fits-all adaptive coil array approach***.

To investigate the potential SNR improvement, a 6-channel receive-only array for 1.5 Tesla was developed. It consists of 4 adaptive channels (see **Figure 3C**) and a rigid 2-channel bottom part (see **Figure 3B**). As reference, a rigid 4-channel reference

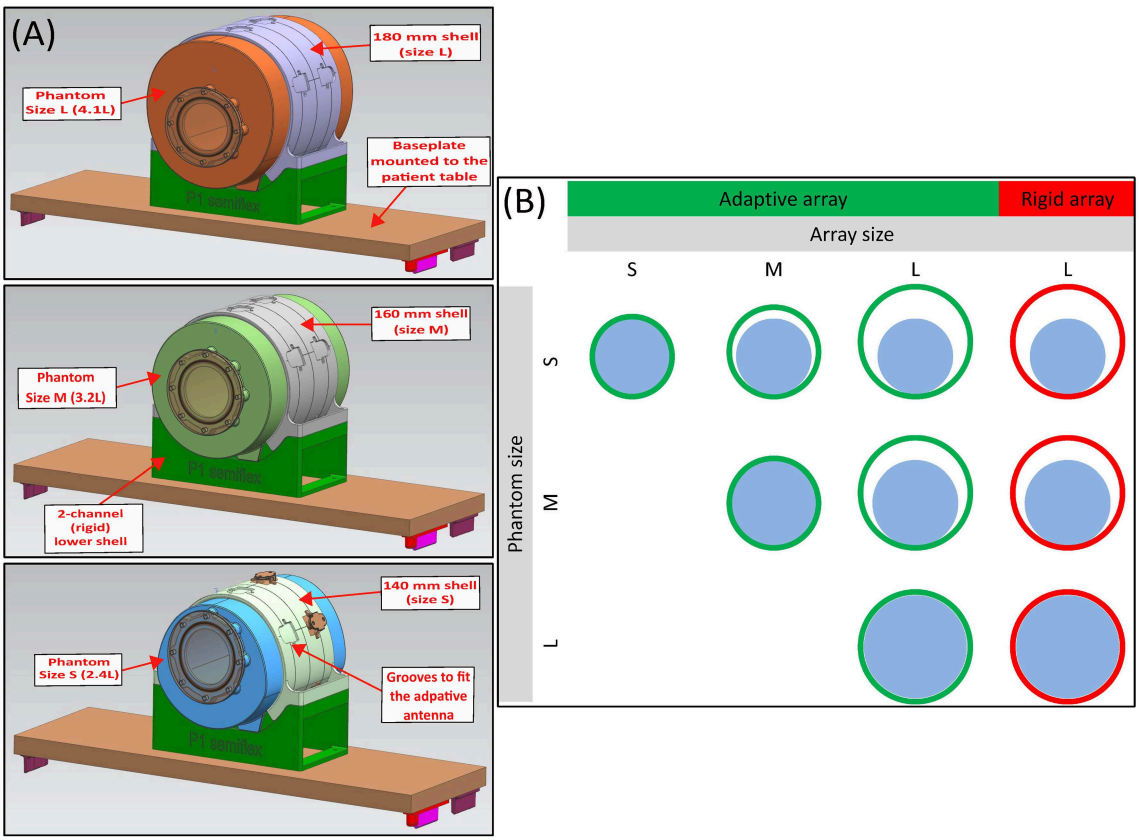


FIGURE 4 | (A) 3D CAD rendering of the adaptive array test setup with 4 stretchable (upper shell) loops and 2 rigid loops in the bottom shell. The adaptive array is evaluated using three differently sized knee phantoms of size L (180 mm diameter, 4.1 l), M (160 mm diameter, 3.2 l), and S (140 mm diameter, 2.4 l). **(B)** To directly compare the 4-channel adaptive array with the four stretchable loops to the 4-channel reference array (each one connect to the 2-channel lower shell upon each measurement), a shell of size 180 mm is 3D laser sintered and the adaptive loops are mounted on it, while the three differently sized knee phantoms (L, M, S) were imaged.

array (see **Figure 3A**) with an inner diameter of 180 mm, the size comparable to a single-row of a commercial 15ch knee coil (Quality Electrodynamics, Mayfield, OH, USA) was constructed, which could also be attached to the rigid 2-channel bottom part. The setup is mounted on a baseplate of the 15-channel knee coil, which is attachable to the patient table (see **Figure 3A**). The four adaptive channels can be attached to three different sized shells with grooves to fit the feed boards, similar to the phantoms (see **Figure 4A**). Geometrical decoupling was pre-adjusted during construction and is maintained with the bottom part for both arrays.

Three differently sized phantoms with 140 mm (size S, 2.4 l), 160 mm (size M, 3.2 l), and 180 mm (size L, 4.1 l) diameter each filled with per 1,000 g H_2O dist.: 1.25 g $NiSO_4 \times 6H_2O$, 5 g NaCl were used for MR imaging. All housing parts including the three phantoms, were 3D laser sintered using the (ObjetEden500V, Stratasys 500V, Rechovot, Israel). The experimental setup can be seen in **Figure 4B**.

All loops were tuned and matched to 63.6 MHz and 50 Ω. The adaptive loops were adjusted at a stretching of 13.34%, which equals the diameter of the phantom size M. Decoupling was $<18\,dB$, pre-amplifier decoupling $>20\,dB$, and matching $<20\,dB$. The stretchable areas of the adaptive array were made of 6 mm wide Amotape® Conduct Elast. #45792. The same properties as evaluated during the material test, were measured again for the knee array in the presence of the other loops, while inductively decoupled.

Finally, SNR was measured for the three phantom sizes using the adaptive and the reference array using the same spin-echo sequence as for the material tests. The smallest configuration of the adaptive array was used to acquire signal and noise images of the phantom size S (140 mm). The 2nd configuration was used to image phantoms S and M and the largest configuration was used to image all three knee phantoms (see **Figure 4B**). The acquired data were exported and reconstructed offline (MATLAB, The Mathworks, Natick, MA, USA).

RESULTS

Material Tests

The goal of the material tests was to identify a suitable material fulfilling the needs of an adaptive coil array. The CuBe2 Strain

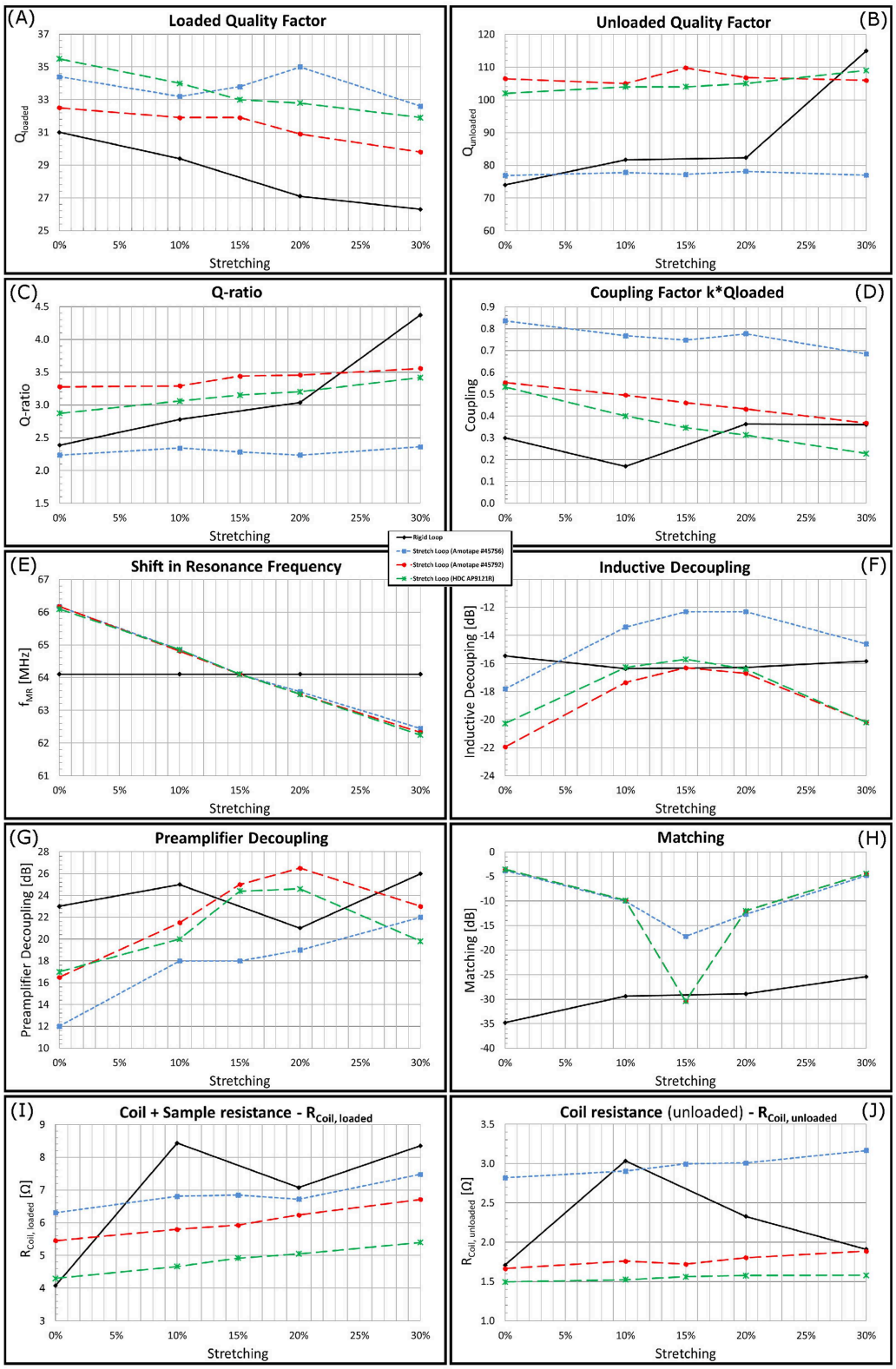
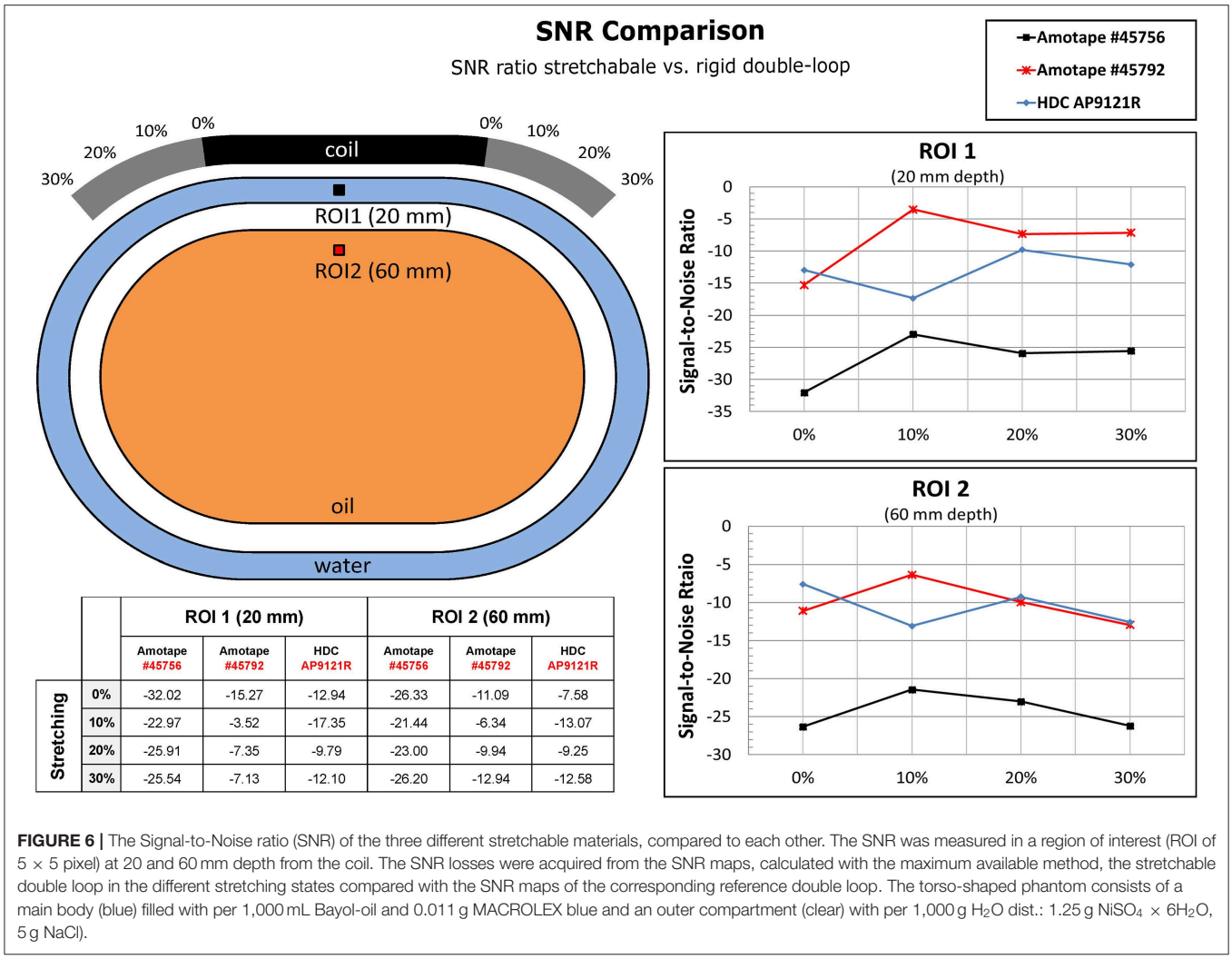


FIGURE 5 | Properties of all tested materials while mounted on the test frame, measured in the laboratory.



Stretching	ROI 1 (20 mm)			ROI 2 (60 mm)		
	Amotape #45756	Amotape #45792	HDC AP9121R	Amotape #45756	Amotape #45792	HDC AP9121R
0%	-32.02	-15.27	-12.94	-26.33	-11.09	-7.58
10%	-22.97	-3.52	-17.35	-21.44	-6.34	-13.07
20%	-25.91	-7.35	-9.79	-23.00	-9.94	-9.25
30%	-25.54	-7.13	-12.10	-26.20	-12.94	-12.58

FIGURE 6 | The Signal-to-Noise ratio (SNR) of the three different stretchable materials, compared to each other. The SNR was measured in a region of interest (ROI of 5 × 5 pixel) at 20 and 60 mm depth from the coil. The SNR losses were acquired from the SNR maps, calculated with the maximum available method, the stretchable double loop in the different stretching states compared with the SNR maps of the corresponding reference double loop. The torso-shaped phantom consists of a main body (blue) filled with per 1,000 mL Bayol-oil and 0.011 g MACROLEX blue and an outer compartment (clear) with per 1,000 g H_2O dist.: 1.25 g $NiSO_4 \times 6H_2O$, 5 g NaCl).

Springs were found to introduce a too high inductance and were, therefore, not further evaluated.

All tested materials were sufficiently stretchable to cover the envisioned element size. Bench measurement results for the rigid double-loop array and the three stretchable double-loop arrays are summarized in **Figure 5** and show the dependence of the following parameters on the amount of stretching: Q_{loaded} (**Figure 5A**), $Q_{unloaded}$ (**Figure 5B**), Q_{ratio} (**Figure 5C**), *coupling coefficient k* (**Figure 5D**), *shift in resonance frequency* (**Figure 5E**), *inductive decoupling* (**Figure 5F**), *pre-amplifier decoupling* (**Figure 5G**), *matching* (**Figure 5H**), *coil+sample resistance (loaded)* (**Figure 5I**), and *unloaded coil resistance* (**Figure 5J**). All coils showed a shift in resonance frequency of about−125 kHz per mm elongation by stretching (see **Figure 5E**). Amotape® #45792 showed consistently highest Q-ratio, best inductive and preamplifier decoupling. The loops with Amotape® #45792 showed coupling coefficients and coil resistance values in the unloaded and loaded cases which lie between the values of the other materials. All measured parameters of the stretchable loops were compared to rigid loops with equivalent sizes. Complete measurement data is listed in the **Figure 11**.

The results of SNR measurements in the MR scanner are shown in **Figure 6**. In most cases and on average, the stretchable loops with Amotape® #45792 showed the least loss in SNR (8/9% in 20/60 mm depth, respectively) as compared to the rigid double coil array.

Adaptive Knee Array

For the adaptive knee array also Amotape® #45792 was used. The measured shift in resonance frequencies for the 4 stretchable loops within the adaptive array when stretched from a knee diameter of 140 to 180 mm, ranged from 66.39 to 61.23 MHz. This corresponds to a stretching of up to 26.68% from the original array size, the stretchable lengths × change from 60.1 to 91.5 mm. The overall size of an adaptive loop varies between 90.6 × 60 mm (un-stretched) and 122 × 60 mm (fully stretched). This resulted in a frequency shift of 164 kHz per mm elongation, which is comparable to the 128 kHz per mm as measured during the material tests for Amotape® #45792.

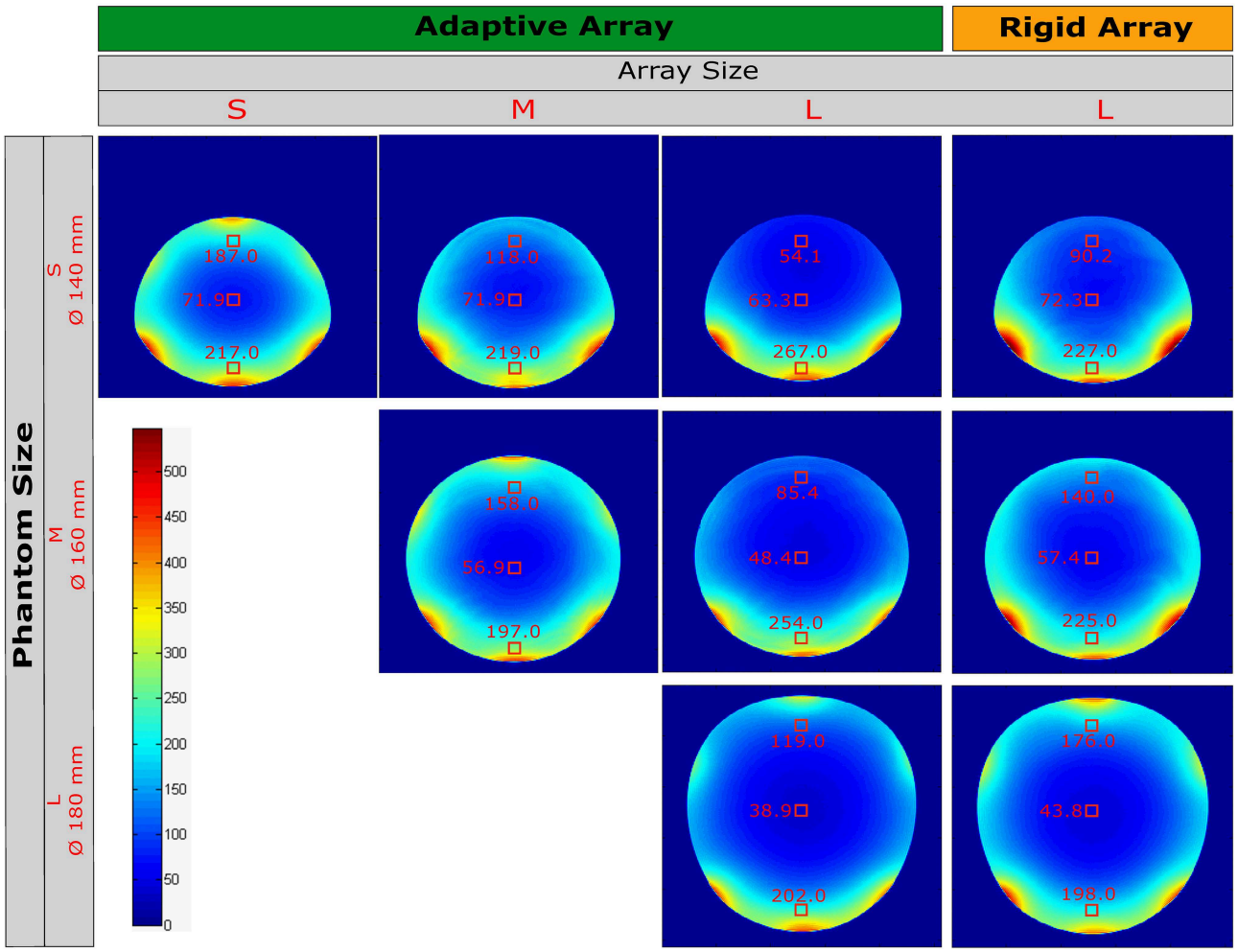


FIGURE 7 | SNR comparison for the adaptive and rigid knee arrays for different phantom sizes. The SNR was measured at three representative points: at the center of each phantom, 25 mm away from the lower shell and 25 mm away from the top of each phantom. The experimental setup is illustrated in **Figure 4B**.

Q-ratios for all loops of both the rigid reference and the adaptive knee array were larger than 2, and therefore in sample noise dominance. The unloaded coil resistances $R_{Coil,unloaded}$ of the individual antenna elements for the 180 mm knee phantom ranged from 0.84 to 1.39 Ω (reference array) and from 0.99 to 2.13 Ω (adaptive array). $R_{Coil,loaded}$ for the same phantom size ranged from 2.38 to 3.67 Ω (reference array) and 2.51 to 4.79 Ω (adaptive array).

The *coupling factor* k^*Q_{loaded} between the loops of the *reference array* ranged from 0.048 to 0.125 and from 0.14 to 1.27 for the *adaptive array*. For the adaptive array this is higher than expected, which can be explained by the higher number of channels and their closer positioning as compared to the setting with 2 loops during the material tests, i.e., each loop was not only influenced by the direct neighbor, but also by all others.

Inductive decoupling ranged between 16.7 and 26.9 dB for the reference array and between 36.8 and 13.9 dB for the adaptive array. The best decoupling in the adaptive array was achieved at maximum stretching, as expected from the results of the material tests.

(Matching) ranged from −43.0 to −18.1 dB for the reference array. The adaptive array adjusted for the size M phantom was matched between −23.4 and −16.7 dB. At minimum and maximum stretching of the adaptive array loops, matching was completely off between −1.6 to −2.4 dB. Same results were achieved with the *pre-amplifier decoupling*. The reference array showed reasonable pre-amplifier decoupling values of 22.0 to 29.8 dB, whereas the loops of the adaptive array, from minimal to maximal stretching of the individual elements, ranged from 12.6 to 26.69 dB. Adjusted to the size M phantom, pre-amplifier decoupling for the adaptive array ranged from 19.6 to 26.5 dB, and is comparable to the values of the reference array.

Figure 7 illustrates the SNR images acquired with the reference and the adaptive array using the three different phantom sizes. SNR values were measured in single voxels in the middle transverse slice at three different depths from the surface, relative to the knee phantom size used. In a direct comparison

between the adaptive and the reference array using phantom size L, the SNR of the adaptive array was worse at 20 mm below the knee phantoms surface (from the top), compared to the reference array. The SNR values in the center are just slightly worse for the adaptive array. When using smaller phantom sizes, the adaptive array, tuned and matched to phantom size M, achieves equal SNR values at the phantom center and much higher SNR values especially below the surfaces of the phantoms. Noise correlation is not exceeding 0.4 in any configuration of the reference or the adaptive array (see **Figure 10**).

In **Figure 8** one can see the SNR comparison between the arrays using the three different knee phantoms. Comparing at phantom size L, it is evident that the adaptive array stretched to sizes S, M and L (but always tuned and matched for array size M and phantom size M) performs worse (−10 to −40%) than the reference array (tuned and matched for array size L and phantom

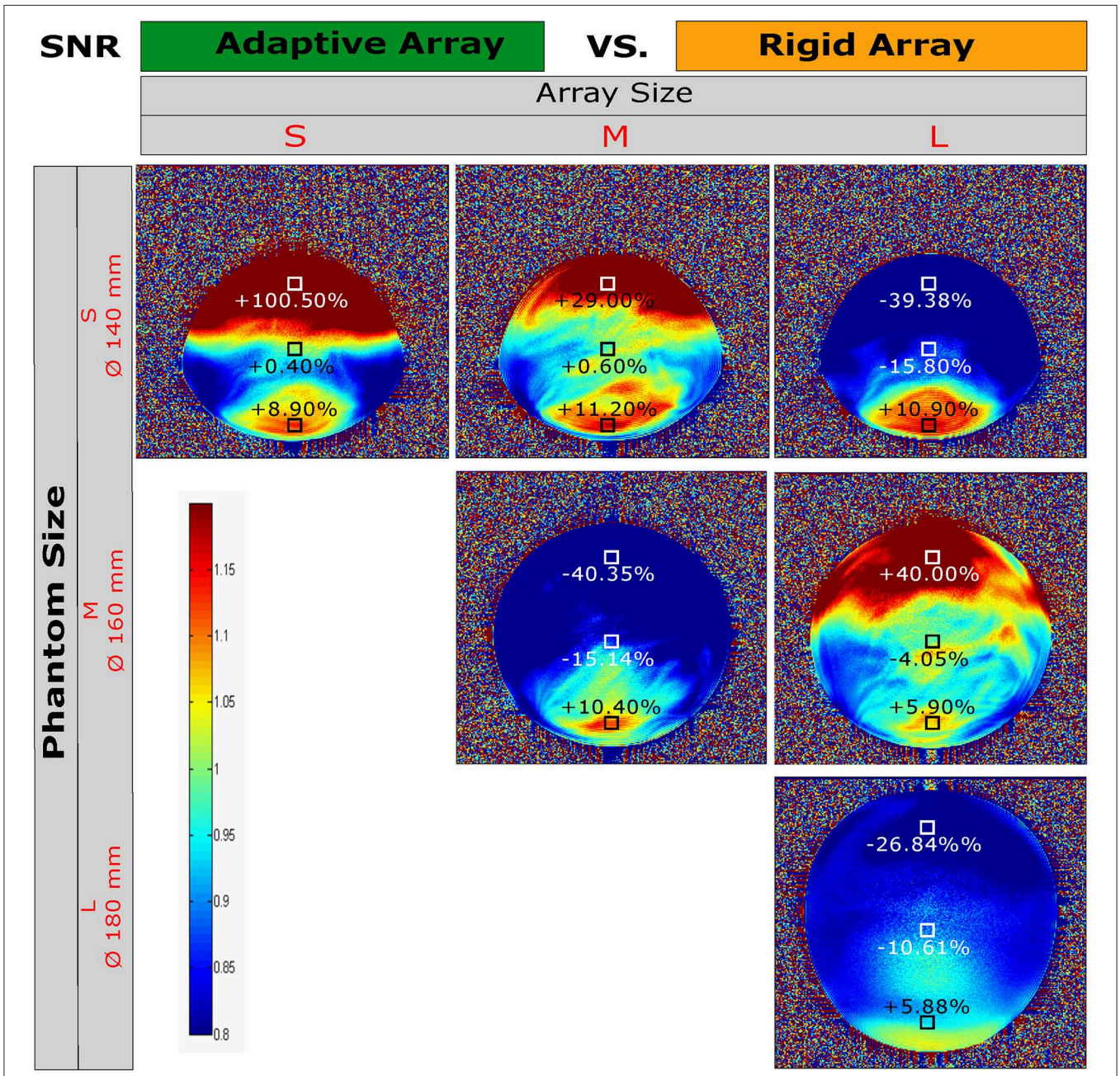


FIGURE 8 | SNR images of the adaptive and the rigid knee array compared to each other using different phantom (knee) sizes. A SNR gain or loss at the ROI is indicated with "+" and "−" symbol and the corresponding percentage value.

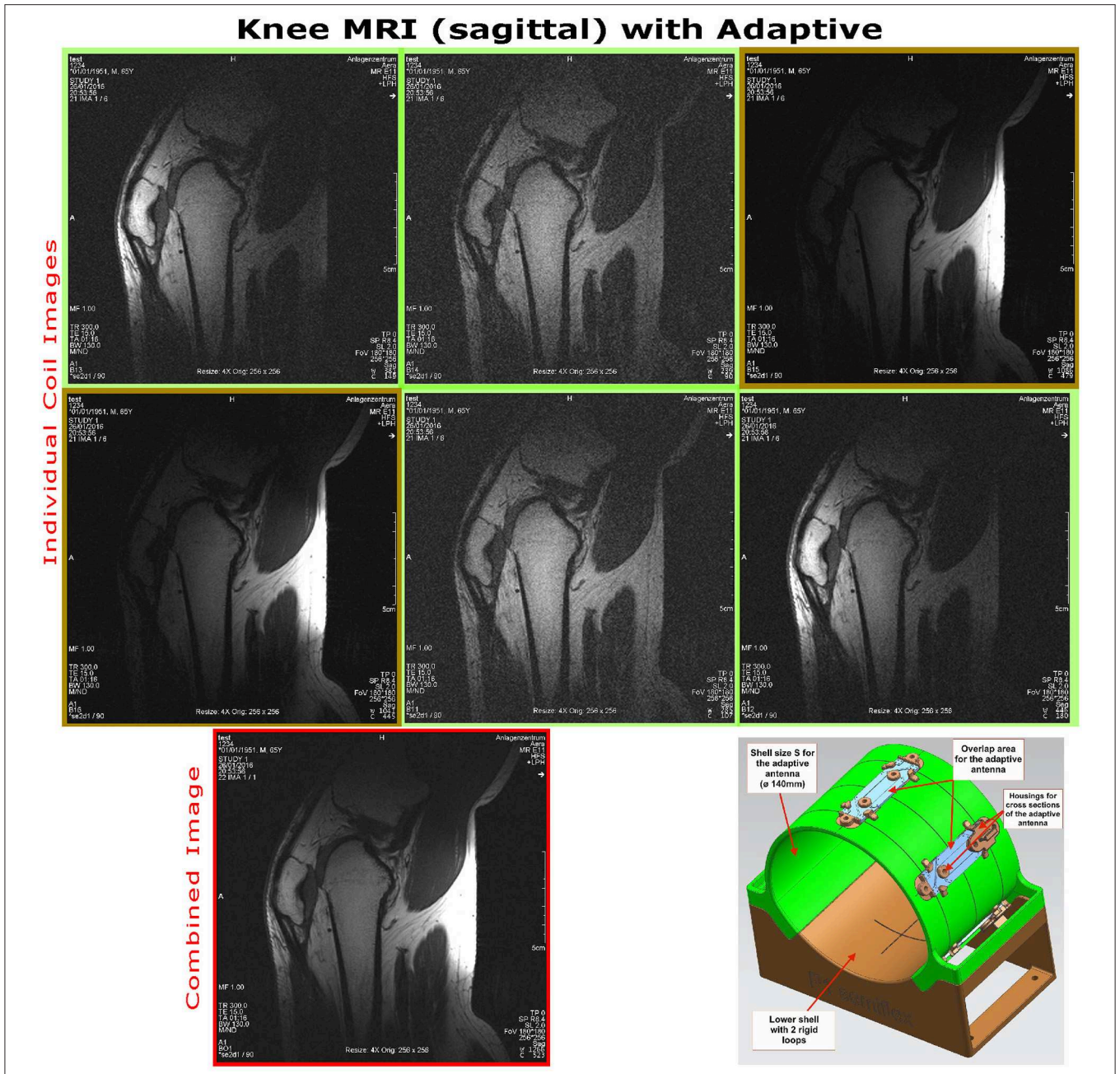


FIGURE 9 | MRI images of the individual channels from the 6ch adaptive array, imaging the right knee of a volunteer. Image 3 (1st row, 3rd column) and 4 (2nd row, 1st column) are from the 2 channels in the lower shell. The image was acquired using a spin-echo sequence in sagittal orientation (TE = 15 ms, TR = 300 ms, FoV = 180 mm, TA = 1:16 min, slice thickness = 2 mm, acq. matrix = 256 × 256, voxel = 1.2 × 1.2mm, bandwidth = 130 Hz/pixel).

size L). However, as expected due to the closer fit to the sample, an SNR gain of 29% (array size M, phantom size S), 40% (array size M, phantom size M), or 100% (array size S, phantom size S) was observed in the voxel near the adaptive part of the arrays. SNR in the center of the phantoms was approximately equal to the rigid array, and a consistent 5–10% SNR gain was found for the voxels near the rigid bottom part.

Figure 9 shows sagittal *in vivo* MR images of the knee center using the adaptive array, displayed as uncombined single channel images, and the combined image. The imaged knee had a very small diameter of 100 mm at the center, even smaller than the smallest size configuration of the adaptive array (130 mm), therefore, the optimal fit of the adaptive array was not reached.

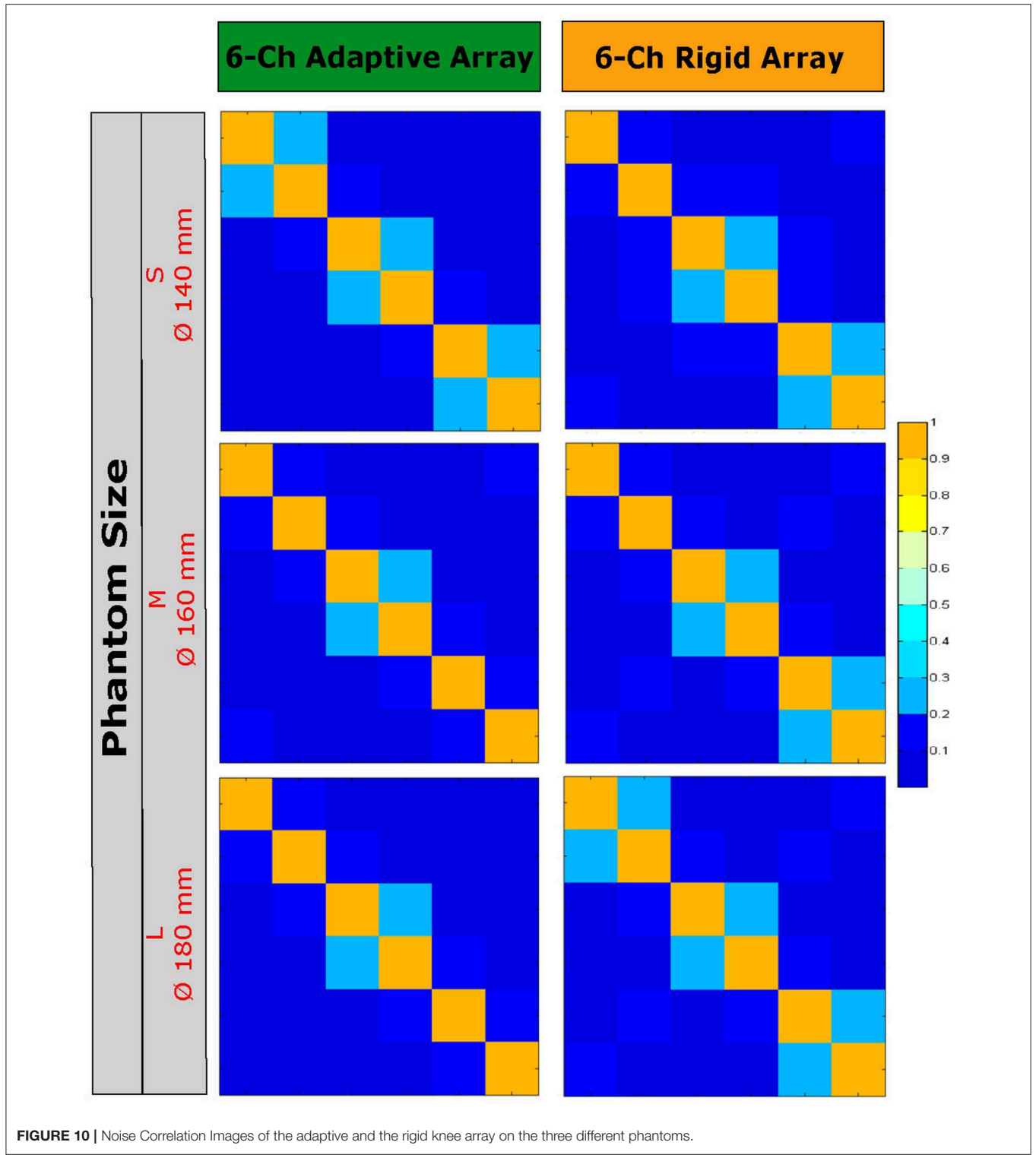


FIGURE 10 | Noise Correlation Images of the adaptive and the rigid knee array on the three different phantoms.

DISCUSSION

An approach for size- and shape-adaptable receive elements using partly-stretchable conductors is presented and its feasibility for *in vivo* MR imaging is demonstrated. Four different stretchable materials were investigated and a material with 19 strands of meandered conductors on an elastic substrate was identified as the best-performing solution in terms of decoupling and achievable SNR. A similar material with only seven strands showed slightly worse performance due to its higher equivalent

Amotape #45756		Shift of f_{MR} [MHz]		Inductive Decoupling [dB]		Coupling factor $k*Q_{loaded}$		Matching [dB]		Preamplifier Decoupling [dB]		Quality Factor Q				Q-ratio		Coil Resistance [ohm]			
		Rigid Loop	Stretch Loop	Rigid Loop	Stretch Loop @ f_{MR}	Rigid Loop	Stretch Loop @ f_{MR}	Rigid Loop	Stretch Loop @ f_{MR}	Rigid Loop	Stretch Loop @ f_{MR}	Rigid Loop		Stretch Loop		Rigid Loop	Stretch Loop @ f_{MR}	Rigid Loop		Stretch Loop @ f_{MR}	
												Q_{loaded}	$Q_{unloaded}$	Q_{loaded}	$Q_{unloaded}$			Q_{loaded}	$Q_{unloaded}$	Q_{loaded}	$Q_{unloaded}$
Stretchable Area	0%	63.60	65.67	-15.46	-17.80	0.300	0.836	-34.80	-3.80	23.00	12.00	31.00	74.00	34.40	76.90	2.39	2.24	4.08	1.71	6.31	2.82
	10%	63.60	64.35	-16.37	-13.40	0.169	0.768	-29.40	-10.00	25.00	18.00	29.40	81.70	33.20	77.80	2.78	2.34	8.43	3.03	6.81	2.90
	15%	–	63.60	–	-12.30	–	0.748	–	-17.20	–	18.00	–	–	33.80	77.20	–	2.28	–	–	6.84	3.00
	20%	63.60	63.07	-16.28	-12.30	0.363	0.777	-28.90	-12.70	21.00	19.00	27.10	82.30	35.00	78.20	3.04	2.23	7.07	2.33	6.72	3.01
	30%	63.60	61.95	-15.83	-14.60	0.360	0.685	-25.40	-4.80	26.00	22.00	26.30	115.00	32.60	77.00	4.37	2.36	8.35	1.91	7.48	3.17

Amotape #45792		Shift of f_{MR} [MHz]		Inductive Decoupling [dB]		Coupling factor $k*Q_{loaded}$		Matching [dB]		Preamplifier Decoupling [dB]		Quality Factor Q				Q-ratio		Coil Resistance [ohm]			
		Rigid Loop	Stretch Loop	Rigid Loop	Stretch Loop @ f_{MR}	Rigid Loop	Stretch Loop @ f_{MR}	Rigid Loop	Stretch Loop @ f_{MR}	Rigid Loop	Stretch Loop @ f_{MR}	Rigid Loop		Stretch Loop		Rigid Loop	Stretch Loop @ f_{MR}	Rigid Loop		Stretch Loop @ f_{MR}	
												Q_{loaded}	$Q_{unloaded}$	Q_{loaded}	$Q_{unloaded}$			Q_{loaded}	$Q_{unloaded}$	Q_{loaded}	$Q_{unloaded}$
Stretchable Area	0%	63.60	65.67	-15.46	-21.94	0.300	0.553	-34.80	-3.44	23.00	16.50	31.00	74.00	32.50	106.50	2.39	3.28	4.08	1.71	5.45	1.66
	10%	63.60	64.31	-16.37	-17.35	0.169	0.496	-29.40	-9.70	25.00	21.50	29.40	81.70	31.90	105.00	2.78	3.29	8.43	3.03	5.79	1.76
	15%	–	63.60	–	-16.30	–	0.461	–	-19.80	–	25.00	–	–	31.90	109.80	–	3.44	–	–	5.92	1.72
	20%	63.60	62.99	-16.28	-16.70	0.363	0.432	-28.90	-11.60	21.00	26.50	27.10	82.30	30.90	106.80	3.04	3.46	7.07	2.33	6.24	1.80
	30%	63.60	61.83	-15.83	-20.18	0.360	0.367	-25.40	-4.13	26.00	23.00	26.30	115.00	29.80	106.00	4.37	3.56	8.35	1.91	6.71	1.89

Highly Ductile Copper #AP9121R		Shift of f_{MR} [MHz]		Inductive Decoupling [dB]		Coupling factor $k*Q_{loaded}$		Matching [dB]		Preamplifier Decoupling [dB]		Quality Factor Q				Q-ratio		Coil Resistance [ohm]			
		Rigid Loop	Stretch Loop	Rigid Loop	Stretch Loop @ f_{MR}	Rigid Loop	Stretch Loop @ f_{MR}	Rigid Loop	Stretch Loop @ f_{MR}	Rigid Loop	Stretch Loop @ f_{MR}	Rigid Loop		Stretch Loop		Rigid Loop	Stretch Loop @ f_{MR}	Rigid Loop		Stretch Loop @ f_{MR}	
												Q_{loaded}	$Q_{unloaded}$	Q_{loaded}	$Q_{unloaded}$			Q_{loaded}	$Q_{unloaded}$	Q_{loaded}	$Q_{unloaded}$
Stretchable Area	0%	63.60	65.59	-15.46	-20.27	0.300	0.533	-34.80	-3.54	23.00	17.00	31.00	74.00	35.50	102.00	2.39	2.87	4.08	1.71	4.30	1.50
	10%	63.60	64.35	-16.37	-16.27	0.169	0.400	-29.40	-9.86	25.00	20.00	29.40	81.70	34.00	104.00	2.78	3.06	8.43	3.03	4.66	1.52
	15%	–	63.60	–	-15.70	–	0.347	–	-30.40	–	24.40	–	–	33.00	104.00	–	3.15	–	–	4.92	1.56
	20%	63.60	62.99	-16.28	-16.40	0.363	0.312	-28.90	-12.00	21.00	24.60	27.10	82.30	32.80	105.00	3.04	3.20	7.07	2.33	5.05	1.58
	30%	63.60	61.75	-15.83	-20.20	0.360	0.228	-25.40	-4.42	26.00	19.80	26.30	115.00	31.90	109.00	4.37	3.42	8.35	1.91	5.40	1.58

FIGURE 11 | All measured values in the laboratory of each double-loop using the materials described in **Table 2** (except *$CuBe_2$* strain springs) with the experimental setup described in **Figure 1D**.

series resistance. A solenoidal spring from CuBe2 was excluded from performance tests since it exhibited a too high inductance which would have required impractically low capacitance values to achieve resonance at the Larmor frequency. In direct comparison of the best stretchable double-loops to rigid reference double-loops, the SNR loss is below 10% on average. A 6-channel knee array prototype with two rigid and four stretchable elements was developed and a thorough comparison to geometrically identical rigid standard loop coil arrays was performed. A considerable SNR gain of up to 100% was demonstrated, which shows that the effect of better conformity to the sample outweighs the SNR penalty for stretchable coils. This penalty arises from the facts that the stretchable coils exhibit inherently higher coil losses and can only be optimized in a single state of stretching in terms of resonance frequency, matching, and decoupling. The resulting variation in impedance by stretching affects the optimum noise matching to the preamplifier, thus degrading SNR.

To minimize these effects, the stretchable loops were tuned and matched, and their geometrical overlap optimized in an average stretching configuration. Evidently, a reduction of the stretchable area would lead to lower frequency shift, but would on the other hand limit the range of patient sizes that could be imaged. Additional techniques to compensate for the change of coil characteristics upon stretching would be beneficial, especially at extreme positions of elongation away from the optimized size. Tuning and matching could be restored by automatic tuning and matching techniques. The approach of using varactor-diodes as voltage-controlled tuning elements to match the impedance of the coil elements has been introduced very early [7, 37]. They were also used to design a closed-loop with automatic tuning and matching circuit for a flexible EPR surface resonator [38], or a microcontroller-based automatic tuning technique for MRI [20]. The drawbacks are that the procedure took about 1 min to complete, and involved a physical disconnection of the local coil from the scanner. A later approach [17] for microcontroller-based automatic tuning of electronics, allowed tuning in the scanner in under 1 s, but can only handle frequency shifts up to 10%,

	AMOTAPE® #45756	AMOTAPE® #45792	Highly Ductile Cooper #AP9121R	Rigid Reference Loops			
				0%	10%	20%	30%
Component	Value [pF]	Value [pF]	Value [pF]	Value [pF]	Value [pF]	Value [pF]	Value [pF]
C_1	56	100	82	82	68	68	100
C_2	27	0	39	56	56	47	12
C_{f1}	68	82	82	100	100	100	82
C_5	0	0	27	47	0	0	0
C_6	23	30	18	43	17	28	20
C_3	82	100	82	82	82	82	150
C_4	100	100	100	100	100	100	100
C_7	470	40	40	100	100	270	470
C_8	330	470	270	150	150	100	0
C_{ges}	10.82	13.24	15.41	19.78	10.09	13.05	11.39

FIGURE 12 | Component values for all built rigid and stretchable loops.

yet still, this technique could be a promising candidate for further investigation.

A possible solution to handle the increased inductive coupling introduced due to the frequency shift when stretching the coil, would be achieved by departing from single coil resonances and rather operate in response plateaus between multiple resonance peaks [16]. The advantage of inductive decoupling is its broadband decoupling effect. A mechanical system introducing the required variation of the overlap area between adjacent elements upon the stretching would possibly maintain good decoupling and improve SNR. However, measurements of the coupling coefficient in this work showed that non-ideal decoupling was not a major concern in this case.

Although sample losses were dominant for the investigated stretchable coils (all $Q_{ratio} > 2$), other stretchable materials that might have better conductivity, like carbon nanotubes in rubber-like stretchable support material or silver/gold antenna structures integrated into Polydimethylsiloxane (PDMS), may also be of interest in the development of stretchable coils.

CONCLUSION

Using array elements with stretchable parts, a viable solution to size- and shape-adaptive coils was demonstrated. Despite a slight SNR loss of 10% in direct comparison to rigid standard loop coils in identical geometrical setup, a considerable SNR gain of up to 100% with an adaptive 6-channel prototype array over a geometrically identical rigid array could be shown in knee phantoms of different sizes. This increase is due to the better form-fitting of the adaptive array to the samples.

This work aims at investigating a novel technology for stretchable and flexible RF coils. To enable the presented methodology for practical application or clinical use, undesirable effects of coil stretching, such as frequency shift, mismatch and imperfect decoupling are yet to be handled.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

BG conceptualized the work, did the calculations, created the design which was later fabricated, and sketched further ideas mentioned in the manuscript. Mechanical design was done by SZ. Experiments were designed by BG and SZ and other members of the Local Coil Laboratory at Siemens Healthcare GmbH in Erlangen (Germany). The manuscript was written by BG and co-authored by EL and reviewed by all authors.

FUNDING

BG had no conflict of interest to declare during the work conducted in 2015 and 2016. The work was partially

funded by the National Institute of Health (project # 1U01EB025162-01) and the anniversary fund of the Austrian National Bank (OeNB) (project # 17980) to EL. RR and SZ are employees of Siemens Healthcare GmbH, Erlangen (Germany).

ACKNOWLEDGMENTS

The authors would like to thank the members at the local coil lab in Erlangen as well as their former leader Hubertus Fischer for the support in this research.

REFERENCES

1. Moser E, Laistler E, Schmitt F, Kontaxis G. Ultra-High Field NMR and MRI - The Role of Magnet Technology to Increase Sensitivity and Specificity. *Front Phys.* (2017) 5:33. doi: 10.3389/fphy.2017.00033
2. Moser E. Ultra-high field magnetic resonance: why and when? *World J Radiol.* (2010) **2**:37–40. doi: 10.4329/wjr.v2.i1.37
3. Keil B, Wald LL. Massively parallel MRI detector arrays. *J Magn Reson.* (2013) **229**:75–89. doi: 10.1016/j.jmr.2013.02.001
4. Ackerman J, Grove T, Wong G, Gadian D, Radda G. Mapping of metabolites in whole animals by 31P NMR using surface coils. *Nat Mag.* (1980) **283**:167–70. doi: 10.1038/283167a0
5. Kneeland JB, Hyde JS. High-resolution MR imaging with local coils. *J Radiol.* (1989) **171**:1–7. doi: 10.1148/radiology.171.1.2648466
6. Roemer PB, Edelstein WA, Hayes CE, Souza SP, Mueller OM. The NMR phased array. *Magn Reson Med.* (1990) **16**:192–225. doi: 10.1002/mrm.1910160203
7. Boskamp E. Improved surface coil imaging in MR: decoupling of the excitation and receiver coils. *J Radiol.* (1985) **157**:449–52. doi: 10.1148/radiology.157.2.4048454
8. Boskamp EB. A new revolution in surface coil technology - the array surface coil. In: *Proceedings of the ISMRM 6th Annual Meeting and Exhibition*. New York, NY (1987).
9. Wiesinger F, De Zanche N, Pruessmann KP. Approaching ultimate SNR with finite coil arrays. In: *Proceedings of the ISMRM 13th Annual Meeting and Exhibition.* Miami Beach, FL (2005). p. 672.
10. Hardy CJ, Cline HE, Giaquinto RO, Niendorf T, Grant AK, Sodickson DK. 32-element receiver coil array for cardiac imaging. *Magn Reson Med.* (2006) **55**:1142–9. doi: 10.1002/mrm.20870
11. Wiggins GC, Triantafyllou C, Potthast A, Reykowski A, Nittka M, Wald LL. 32-channel 3 Tesla receive-only phased-array head coil with soccer-ball element geometry. *Magn Reson Med.* (2006) **56**:216–23. doi: 10.1002/mrm.20925
12. Schmitt M, Potthast A, Sosnovik DE, Polimeni JR, Wiggins GC, Triantafyllou C, et al. A 128-channel receive-only cardiac coil for highly accelerated cardiac MRI at 3 Tesla. *Magn Reson Med.* (2008) **59**:1431–9. doi: 10.1002/mrm.21598
13. Wiggins GC, Polimeni JR, Potthast A, Schmitt M, Alagappan V, Wald LL. A 96-channel receive-only head coil for 3 Tesla: design optimization and evaluation. *Magn Reson Med.* (2009) **62**:754–62. doi: 10.1002/mrm.22028
14. Vaughan JT, Griffiths JR. *RF Coils for MRI - Encyclopedia of Magnetic Resonance (EMR) Handbooks*. Chichester, UK: John Wiley and Sons Ltd. (2012). p. 334–5.
15. Wright SM. Receiver loop arrays. *Encyclopedia Magn Reson.* (2011) 1–13. doi: 10.1002/9780470034590.emrstm1129
16. Vester M, Biber S, Rehner R, Wiggins GC, Brown R, Sodickson DK. Mitigation of inductive coupling in array coils by wideband port matching. In: *Proceedings of the ISMRM 20th Annual Meeting and Exhibition, Vol. 20.* Melbourne, VIC (2012). p. 2690.
17. Venook RD, Hargreaves BA, Gold GE, Conolly SM, Scott GC. Automatic tuning of flexible interventional RF receiver coils. *Magn Reson Med.* (2005) **54**:983–93. doi: 10.1002/mrm.20616
18. Sohn SM, Gopinath A, Vaughan JT. Electrically auto-tuned RF coil design. In: *Proceedings of the ISMRM 19th Annual Meeting and Exhibition.* Montreal, QC (2011). p. 3826.
19. Sohn SM, DelaBarre L, Gopinath A, Vaughan JT. RF coil design with automatic tuning and matching. In: *Proceedings of the ISMRM 21th Annual Meeting and Exhibition.* Salt Lake City, UT (2013). p. 0731.
20. Rousseau J, Lecouffe P. A new, fully versatile surface coil for MRI. *Magn Reson Med.* (1990) **8**:517–23. doi: 10.1016/0730-725X(90)90061-6
21. Malko JA, McClees EC, Braun IF, Davis PC, Hoffman JC Jr. A flexible Mercury-filled surface coil for MR Imaging. *Am J Neuroradiol.* (1986) 7:6–247.
22. Adriany G, Van de Moortele PF, Ritter J, Moeller S, Auerbach EJ, Akgun C, et al. A geometrically adjustable 16-channel transmit/receive transmission line array for improved RF efficiency and parallel imaging performance at 7 Tesla. *Magn Reson Med.* (2008) 59:590–7. doi: 10.1002/mrm.21488
23. Nordmeyer-Massner JA, De Zanche N, Pruessmann KP. Mechanically adjustable coil array for wrist MRI. *Magn Reson Med.* (2009) **61**:429–38. doi: 10.1002/mrm.21868
24. Port A, Reber J, Vogt C, Marjanovic J, Sporrer B, Wu L, et al. Towards wearable MR detection: a stretchable wrist array with on-body digitization. In: *Proceedings of the ISMRM 26th Annual Meeting and Exhibition.* Paris (2018). p. 17.
25. Corea JR, Flynn AM, Lechene B, Scott G, Reed GD, Shin PJ, et al. Screen-printed flexible MRI receive coils. *Nat Commun.* (2016) **7**:10839. doi: 10.1038/ncomms10839
26. Frass-Kriegl R, de Lara LIN, Pichler M, Sieg J, Moser E, Windischberger C, et al. Flexible 23-channel coil array for high-resolution magnetic resonance imaging at 3 Tesla. *PLoS ONE.* (2018) **13**:e0206963. doi: 10.1371/journal.pone.0206963
27. Hosseinnezhadian S, Frass-Kriegl R, Goluch-Roat S, Pichler M, Sieg J, V'it M, et al. A flexible 12-channel transceiver array of transmission line resonators for 7T MRI. *J Magn Reson.* (2018) **296**:47–59. doi: 10.1016/j.jmr.2018.08.013
28. Zhang B, Sodickson DK, Cloos MA. A high-impedance detector-array glove for magnetic resonance imaging of the hand. *Nat Biomed Eng.* (2018) **2**:570–7. doi: 10.1038/s41551-018-0233-y
29. Ruytenberg T, Webb A, Zivkovic I. Shielded-coaxial-cable coils as receive and transceive array elements for 7T human MRI. *Magn Reson Med.* (2019) **83**:1135–46. doi: 10.1002/mrm.27964
30. Stormont RS, Lindsay SA, Taracila V, Mustafa G, Malik NM, Robb FJL, et al. *Systems for a Radio Frequency Coil for MR Imaging US20190277926A1*. (2019).
31. Gruber B, Zink S. Anatomically adaptive local coils for MRI imaging–evaluation of stretchable antennas at 1.5T. In: *Proceedings of the ISMRM 24th Annual Meeting and Exhibition.* Singapore (2016). p.543.
32. Gruber B, Zink S. Anatomically adaptive coils for mr imaging – a 6-channel demonstrator array study at 1.5 Tesla. In: *Proceedings of the ISMRM 26th Annual Meeting and Exhibition*. Paris. (2018). p. 4289.
33. Gruber B, Jahns K, Zink S. *Adaptive MR Local Coil, US2017089991A1, EP3151025A3, CN107064837A*. (2019).
34. Hong J-S, Lancaster MJ. Couplings of microstrip square open-loop resonators for cross- coupled planar microwave filters. *IEEE Trans Microw Theory Tech.* (1996) **44**:2099–109. doi: 10.1109/22.543968
35. Vizmuller P. *RF Design Guide: Systems, Circuits, and Equations, Series: Artech House Antennas and Propagation Library* (Artech Artech House). (1995) p. 296.

36. Larsson EG, Erdogmus D, Yan R, Principe JC, Fitzsimmons J. SNR-optimality of sum-of-squares reconstruction for phased-array magnetic resonance imaging. *J Magn Reson.* (2003) **163**:121–3. doi: 10.1016/S1090-7807(03)00132-0
37. Doornbos J, Grimbergen H, Booijen P, Strake L, Bloem J, Vielvoye G, et al. Application of anatomically shaped surface coils in MRI at 0.5T. *Magn Reson Med.* (1986) 3:270–81. doi: 10.1002/mrm.1910030210
38. Hirata H, Walczak T, Swartz H. Electronically tuneable surface-coil-type resonator for l-band EPR spectroscopy. *Magn Reson Med.* (2000) **142**:159–67. doi: 10.1006/jmre.1999.1927

Conflict of Interest: BG and EL declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. RR and SZ declare to be employee of Siemens Healthcare GmbH, Erlangen, Germany.

frontiers in Physics

ORIGINAL RESEARCH
published: 15 January 2020
doi: 10.3389/fphy.2019.00237

Multi-Loop Radio Frequency Coil Elements for Magnetic Resonance Imaging: Theory, Simulation, and Experimental Investigation

*Roberta Frass-Kriegl[1], Sajad Hosseinnezhadian[2], Marie Poirier-Quinot[2], Elmar Laistler[1] and Jean-Christophe Ginefri[2]**

[1] Division MR Physics, Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Vienna, Austria, [2] IR4M (Imagerie par Résonance Magnétique et Multi-Modalités), UMR 8081, Université Paris-Sud/CNRS, Université Paris-Saclay, Orsay, France

Magnetic resonance imaging (MRI) is a major imaging modality, giving access to anatomical and functional information with high diagnostic value. To achieve high-quality images, optimization of the radio-frequency coil that detects the MR signal is of utmost importance. A widely applied strategy is to use arrays of small coils in parallel on MR scanners equipped with multiple receive channels that achieve high local detection sensitivity over an extended lateral coverage while allowing for accelerated acquisition and SNR optimization by proper signal weighting of the channels. However, the development of high-density coil arrays gives rise to several challenges due to the increased complexity with respect to mutual decoupling as well as electronic circuitry required for coil interfacing. In this work, we investigate a novel single-element coil design composed of small loops in series, referred to as "multi-loop coil (MLC)." The MLC concept exploits the high sensitivity of small coils while reducing sample induced noise together with an extended field of view, similar to arrays. The expected sensitivity improvement using the MLC principle is first roughly estimated using analytical formulae. The proof of concept is then established through fullwave 3D electromagnetic simulations and validated by B_1 mapping in MR experiments on phantom. Investigations were performed using two MLCs, each composed of 19 loops, targeting MRI at high (3 T) and at ultra-high field strength (7 T). The 3 T and 7 T MLCs have an overall diameter of 12 and 6 cm, respectively. For all investigated MLCs, we demonstrate a sensitivity improvement as compared to single loop coils. For small distances inside the sample, i.e., close to the coil, a sensitivity gain by a factor between 2 and 4 was obtained experimentally depending on the set-up. Further away inside the sample, the performance of MLCs is comparable to single loop coils. The MLC principle brings additional degrees of freedom for coil design and sensitivity optimization and appears advantageous for the development of single coils but also individual elements of arrays, especially for applications with a larger area and shallow target depth, such as skin imaging or high-resolution MRI of brain slices.

Keywords: magnetic resonance imaging, radio frequency coil, surface coil, electromagnetic simulation, B_1 mapping

OPEN ACCESS

Edited by:
Simo Saarakkala, University of Oulu, Finland

Reviewed by:
Stephan Orzada, University of Duisburg-Essen, Germany
Andre Kuehne, MRI.TOOLS GmbH, Germany

****Correspondence:***
Jean-Christophe Ginefri
jean-christophe.ginefri@u-psud.fr

Specialty section:
This article was submitted to Medical Physics and Imaging, a section of the journal Frontiers in Physics

Received: *27 September 2019*
Accepted: *16 December 2019*
Published: *15 January 2020*

Citation:
Frass-Kriegl R, Hosseinnezhadian S, Poirier-Quinot M, Laistler E and Ginefri J-C (2020) Multi-Loop Radio Frequency Coil Elements for Magnetic Resonance Imaging: Theory, Simulation, and Experimental Investigation. Front. Phys. 7:237. doi: 10.3389/fphy.2019.00237

INTRODUCTION

Radio frequency (RF) coils are the front end of the instrumental chain of a magnetic resonance imaging (MRI) system. They are used to generate the RF magnetic field that excites the nuclear spins, and to detect the MR signal, i.e., the RF signal induced by the rotating nuclear magnetization during relaxation. Consequently, RF coils play a major role in MRI since they are the link between the scanner and the sample to be imaged. In order to perform MRI with high diagnostic value, i.e., with high spatial resolution and high signal to noise ratio (SNR), it is of utmost importance to optimize the sensitivity of the RF coil with respect to both, the targeted clinical application and the MRI set-up.

The sensitivity factor of the RF coil quantifies the contribution of the coil to the overall SNR and represents its efficiency to detect the MR signal while minimizing the noise involved in the MR experiment [1]. The two main noise sources are the noise of the coil itself and the noise induced in the coil by the sample [2]. In most clinical MRI applications (typical field strength ≥ 1.5 T), targeting large anatomical sites (e.g., brain, knee) and employing large RF coils (i.e., diameters of several cm), the sample noise largely dominates over the internal coil noise, and is therefore the limiting factor for achieving high detection sensitivity.

The earliest and most often pursued solution to improve the RF coil detection sensitivity is to reduce the coil size, i.e., to use small surface coils [3–5]. This increases the magnetic coupling between the coil and the sample, thus increasing the amplitude of the detected MR signal. In addition, the equivalent volume of sample seen by the coil is reduced and, therefore also the sample induced noise. However, the limited field of view (FoV) of small coils reduces the accessible region of interest (ROI) and may be problematic for applications targeting anatomical regions that extend over an area that is large compared to the target depth.

To overcome this, a widely applied strategy is to use arrays of small coils together with MR scanners featuring multiple receive channels [6–8]. Arrays benefit from the high local detection sensitivity of small surface coils while achieving a lateral coverage comparable to large coils, and are now used in numerous clinical applications of MRI [9]. However, the development of high-density coil arrays evokes additional technological challenges [9, 10] due to the increased complexity with respect to mutual coupling between coils and electronic circuitry required for coil interfacing. Especially the realization of arrays with very small elements becomes impractical either due to fabrication issues with the coil elements themselves, or due to space requirements for the interface components and preamplifiers. On top of that, high-density arrays entail a significant increase of cost due to the high amount of required electronic components and the need for a high number of acquisition channels at the MR scanner.

In this work, we investigate a novel coil design that aims at achieving some benefits of coil arrays as compared to large single loop coils (SLCs), i.e., reduce the sample induced noise by using small coils and achieve a large FoV. Although this novel design doesn't represent an actual alternative to coil arrays, since it neither allows for parallel imaging nor SNR optimization by combining the signals of individual channels with optimal weights, it aims at a comparable sensitivity gain in comparison to large SLCs while relaxing constraints in terms of complexity and cost.

The general concept of this work is to investigate single coil elements composed of small loops in series, in contrast to the coil array principle where the coils are independent and operated in parallel. The association of small loops in series results in a single coil element composed of multiple loops, subsequently referred to as "multi-loop coil (MLC)." MLCs appear particularly advantageous for reducing sample-induced noise that varies with the loop radius to the power of three, while the equivalent noise voltages induced in each of the loops are summed linearly since they are in series. The use of small loops in series may also improve the magnetic coupling between the coil and the sample because the magnitude of the detected MR signal is inversely proportional to the loop radius. Consequently, a significant improvement in detection sensitivity is expected by using MLCs.

Few works reported the use of small loops associated to larger coils, either employing various sized small loops in series to reinforce and homogenize the magnetic coupling of the coil to the sample [11] or employing small loops of the same size equally distributed around a large loop [12, 13], with connections between small loops being alternately reversed so that the magnetic field is in phase. While these two investigations are supported by the same conceptual consideration as the present work, their targets are different, and they face several limitations regarding the freedom for loops positioning and number. Also, in both cases, a large loop is used, which counterbalances the benefit of using small loops and sets a limit to the achievable sensitivity improvement. In addition, an investigation on the effective improvement of the overall detection sensitivity has not been performed so far, neither by simulation nor by MRI experiments.

In this paper, we introduce the theoretical background supporting the MLC principle involving equations for sample and coil losses as well as for the magnetic field produced per unit of current. We present an electromagnetic simulation study to evaluate the MR performance of MLCs and, in particular, the improvement in terms of transmit efficiency, i.e., the magnetic field produced per square root of input power. We show experimental results obtained by MRI, i.e., maps of the transmit efficiency that aim at validating the simulation results and at experimentally demonstrating the sensitivity improvement achieved by MLCs as compared to SLCs.

THEORY

RF Coil Sensitivity

An RF coil can be modeled as a resonant RLC circuit (resistance R, inductance L, capacitance C) tuned to the Larmor frequency of interest and matched to the input impedance of the MR scanner, i.e., typically 50 Ohm. The sensitivity factor of the RF coil S_{RF}, which represents the contribution of the coil to the overall SNR, is the ratio of the induction coefficient, defined as the magnetic field, B_1, per unit current, I, produced by the coil and the equivalent noise voltage associated to the losses involved

in an MRI experiment.

$$S_{\mathrm{RF}} = \frac{\frac{B_1}{I}}{\sqrt{R_{\mathrm{eq}} T_{\mathrm{eq}}}} \tag{1}$$

where $R_{\mathrm{eq}} T_{\mathrm{eq}}$ is the sum of the temperature-weighted resistances associated to different dissipative media and loss mechanisms.

The complete electromagnetic approach to derive expressions for B_1 and losses is given in the literature (see for example [14] p. 127 for losses and p. 206 for B_1/I). Starting from the AC source-current in the coil, one can derive a vector potential. From this the currents induced in the media can be determined, which in general have two components, the eddy current depending on conductivity, and the displacement current depending on permittivity. Using Maxwell-Ampère's equation, one can calculate the magnetic field inside the media originating foremost from the source-current but also from the currents induced in the media. The integral over the real part of the induced currents divided by the conductivity of the media provides the corresponding power loss density. Finally, using Ohm's law, the equivalent resistance is obtained from the power loss density and the current in the coil.

This approach leads to complex integral equations that cannot be solved analytically and require the use of advanced electromagnetic solvers. However, simplified formulae for B_1 and losses neglecting propagation effects have been proposed (see below). Thus, B_1 can be calculated using Biot-Savart's law, neglecting losses due to displacement currents. This approximation is valid if the dimensions of the ROI are small compared to the operating wavelength. For surface coils, when the depth of the targeted ROI does not exceed 10 cm, this assumption is typically fulfilled for field strengths up to 3 T.

Induction Coefficient

The induction coefficient of the coil, $\frac{B_1}{I}$, is the magnetic coupling efficiency of the coil to the sample and is representative of the detected amount of MR signal. $\frac{B_1}{I}$, depends on the area through which the nuclear magnetic flux passes and is set by the winding shape of the coil only.

For a single-turn circular coil of radius a, the theoretical expression for $\frac{B_1}{I}$ in Cartesian coordinates is given in [15]:

$$\frac{B_{1x}}{I} = \frac{\mu_0 xz}{2\pi\alpha^2\beta X^2}\left[\left(a^2+Y^2\right)\mathrm{E}\left(k^2\right) - \alpha^2\mathrm{K}\left(k^2\right)\right] \tag{2}$$

$$\frac{B_{1y}}{I} = \frac{\mu_0 yz}{2\pi\alpha^2\beta X^2}\left[\left(a^2+Y^2\right)\mathrm{E}\left(k^2\right) - \alpha^2\mathrm{K}\left(k^2\right)\right] \tag{3}$$

$$\frac{B_{1z}}{I} = \frac{\mu_0}{2\pi\alpha^2\beta}\left[\left(a^2-Y^2\right)\mathrm{E}\left(k^2\right) + \alpha^2\mathrm{K}\left(k^2\right)\right] \tag{4}$$

with the vacuum permeability μ_0, standard elliptical integrals $\mathrm{E}(k^2)$ and $\mathrm{K}(k^2)$, and the following correspondences:

$$X^2 = x^2 + y^2 \tag{5}$$

$$Y^2 = x^2 + y^2 + z^2 \tag{6}$$

$$\alpha^2 = a^2 + Y^2 - 2aX \tag{7}$$

$$\beta^2 = a^2 + Y^2 + 2aX \tag{8}$$

$$k^2 = 1 - \frac{\alpha^2}{\beta^2} \tag{9}$$

$$\gamma = x^2 - y^2 \tag{10}$$

Along the coil axis, z, the above equations simplify, and the induction coefficient of the coil is expressed as:

$$\frac{B_{1z}}{I} = \frac{\mu_0 a^2}{2\pi\left(a^2+z^2\right)^{\frac{3}{2}}} \tag{11}$$

For distances that are small compared to the coil radius, it varies roughly as a^{-1}, explaining why small surface coils perform better than large ones when investigating ROIs located at the surface of the body. Further away from the coil, i.e., for distances large compared to the coil radius, the induction coefficient varies as a^2 which disfavors smaller coils. However, this can be counterbalanced by combining several small coils operating constructively, since the total induction coefficient is the vector sum of the individual induction coefficients and can tend to equalize that of a large coil at long distance.

Dominant and Non-dominant Noise Sources in MRI

The equivalent noise voltage illustrates the total energy dissipated during the MR experiment. It is proportional to the square-root of the sum of equivalent temperature-weighted resistances according to respective dissipation rates and local temperatures in the different media, $R_{\mathrm{eq}} T_{\mathrm{eq}}$. Several noise mechanisms may be involved in MR experiments and a quantitative comparison of their respective contribution to the overall noise has to be done to identify suitable strategies to improve the RF sensitivity of the coil. The two main noise mechanisms to be considered in current biomedical applications of MRI are the noise of the coil itself and the noise induced in the coil by the sample, i.e., magnetically coupled sample noise [2].

In addition, several, usually non-dominant, noise mechanisms can potentially be involved in MR experiments. For instance, capacitively-coupled sample noise, which tends to be more significant at high field strength [16], can be reduced to a negligible level by using distributed tuning capacitors and inductive coupling transformers [17]. Other noise mechanisms, such as the spin noise [18] and the radiation noise [19], are of marginal relevance for current clinical MRI applications, i.e., below 300 MHz (i.e., ^{1}H Larmor frequency at 7 T). Lastly, the use of electronic components and active devices may introduce additional losses, but they are usually minimized by optimizing the circuit design, e.g., placing high-gain low-noise preamplifiers as close as possible to the coil port.

Magnetically Coupled Sample Noise

Magnetically coupled sample noise is strongly related to the coupling efficiency of the coil, as the corresponding noise voltage is induced in the coil via the same physical pathway as the MR signal. Neglecting displacement currents as explained above, and considering again a circular loop of radius a placed at a distance s to a semi-infinite conducting sample with conductivity

σ, the following approximate formula [20] of the sample induced resistance R_S can be used in many practical situations:

$$R_S = \frac{2}{3\pi}\sigma {\mu_0}^2\omega^2 a^3 \arctan\left(\frac{\pi a}{8s}\right) \quad (12)$$

Besides the quadratic dependence of R_S on the operating frequency, indicating that these losses may be predominant over other losses at high field strength, it can be observed that R_S varies as the power of 3 with the coil radius. This explains the large advantage of using small coils when sample noise dominates. It can be noticed that for coils with a radius that is small compared to the distance between the coil and the sample, the arctan-function varies as *a,* and the sample induced resistance then depends on the coil radius as the power 4. This consideration is the basis for the pinpoint coil concept that showed the potential improvement in sensitivity by using very small coil when the sample noise dominates, even for imaging regions located deep inside the sample [21].

When using several coils, in parallel as in arrays, or in series as in MLCs, the overall sample induced noise is larger than the sum of the sample induced noise in all individual loops. This noise increase is due to electrical coupling between the coils via the sample, referred to as noise correlation [6]. The analytical formulae to calculate the mutual resistance of several coils implies the determination of spatial distribution of the electric fields produced by all coils, and requires advanced computations [22] that are beyond the scope of the presented analytical evaluation of MLCs. Alternatively, one can consider a relative increase of the sample induced noise based on the electrical coupling coefficient as a function of the distance between loops [6]. The electric coupling coefficient between two coils can be defined in analogy to magnetic coupling:

$$k_{e(ik)} = \frac{R_{Sik}}{\sqrt{R_{Si}R_{Sk}}} \quad (13)$$

Where R_{Sik} is the mutual resistance between the coil i and coil k and R_{Si}, R_{Sk} are the sample induced resistance in the coils when isolated, i.e., without noise correlation, given by Equation (12). In the MLC design, all loops have the same sample induced resistance when large enough phantoms are employed, i.e., $R_{Si} = R_{Sk}$. The value of the total sample resistance is then the sample induced resistance of each loop isolated plus twice the mutual resistance between loops:

$$R_{S\,TOT} = \sum_{i=1}^{N} R_S + 2\sum_{i=1}^{N}\sum_{\substack{k=1\\k\neq i}}^{N} R_{Sik} \quad (14)$$

Coil Noise

The standard technology to fabricate RF coils employs wound conducting wires intersected by lumped element capacitors. In this case, the total coil losses account for the ohmic losses of the winding as well as the real losses of the capacitors used to tune the coil.

The general formula for the ohmic resistance of a conducting loop of radius a made of a wire of radius r, with electrical resistivity ρ, and skin-depth $\delta = \sqrt{2\rho/(\mu_0\omega)}$, can be approximated by the following expression [23] when operating at frequencies high enough so that $\delta << r$:

$$R_{Cwire} = \frac{\rho a}{r\delta} = \frac{a}{r}\sqrt{\frac{\rho\mu_0\omega}{2}} \quad (15)$$

In the case of a conducting loop of outer radius a made of a flat conducting strip of width w the above expression becomes [23]:

$$R_{Cstrip} = \frac{\rho\pi a}{2w\delta} = \frac{\pi a}{2w}\sqrt{\frac{\rho\mu_0\omega}{2}} \quad (16)$$

This estimation only considers the "classical" skin effect and neglects the lateral skin effect contributions for flat strip conductors, which tend to dominate especially at higher frequencies [24–26]. In addition, the conductor losses of multi-turn and multi-loop coils can be increased due to the proximity effect [27], which constrains the current distribution to an even smaller region than the skin effect. This loss contribution depends on the exact coil design and is neglected for the following rough loss estimation.

Capacitor losses originate from two different phenomena with relative contributions depending on the operating frequency. The first component are dielectric losses inside the capacitor material itself characterized by the loss tangent, tanδ ([14] p. 127). Second, the metallic parts of the capacitors, such as contacts for soldering, generate metallic losses, similarly to conductors. The total losses of capacitors, referred to as the equivalent series resistance (ESR) are the sum of the dielectric and metallic resistances, and are usually available from data sheets. Regarding the typical operating frequency range of RF coils in MRI, and accounting for the conductivity and permittivity of typical capacitor materials, metallic losses dominate over dielectric losses. In addition to the capacitor losses, losses associated to solder joints used to mount the capacitors onto the coil winding should be considered as it may have a significant contribution to the overall noise [23].

One can notice that when employing other technologies than that of lumped components to fabricate the coil, the two above mentioned noise sources are still of concern. Considering the case of self-resonant coils based on the transmission line principle (e.g., [28]), losses within the substrate, similarly to capacitors losses, may contribute to the overall coil noise and should be considered. This is particularly of concern when using low quality substrates, such as FR4 or Polyimide (e.g., Kapton®) whose loss tangent limits the achievable overall quality factor of the coil. However, by choosing materials with a low loss tangent, such as Polytetrafluoroethylene (PTFE), dielectric losses within the substrate can be kept at a negligible level compared to the ohmic losses of the coil conductor.

Rough Estimate of the Sensitivity Improvement Expected With MLCs

As discussed above, reducing the coil size can provide a significant reduction of the sample induced losses together with

an increase of the induction coefficient. In this section, we roughly evaluate the theoretical gain in sensitivity expected by using an MLC as compared to an SLC achieving an equivalent FoV. In this regard, we will consider an SLC of radius A and compare its losses and induction coefficient to those of an MLC composed of N small loops of radius $a = \frac{A}{\sqrt{N}}$ connected in series.

Estimation of the Induction Coefficient

Considering sample-to-coil distances small compared to the coil radius, which corresponds to the usual case when using surface coils, the induction coefficient is inversely proportional to the coil radius (see Equation 11). Consequently, a small loop of radius $a = \frac{A}{\sqrt{N}}$, will produce a B_1/I at close distance along its axis that is $\sqrt{N}$-times higher than that of a larger loop of radius A. It can be assumed that in the case of N small loops in series, close to one small loop the other loops will not produce a significant B_1 along its axis. So, at close distance from the loops, only a benefit from the size reduction can be expected using the MLC principle. Further away from the loop, B_1/I on its axis varies as the square of the radius. In this case, N small loops of radius a, are expected to produce a B_1/I that is comparable to that of a larger loop of radius A. It finally appears from this rough estimate that MLCs achieve an induction coefficient higher or equal to that of a large coil.

Estimation of Sample Losses

When considering the sample resistance dependency on the coil radius, as shown in Equation (12), and for sample to coil distances small as compared to the coil radius, it appears that a large loop of radius A will result in a sample-induced resistance proportional to A^3, whereas N small loops of radius $a = \frac{A}{\sqrt{N}}$, will result at a first glance in a total sample-induced resistance proportional to $\frac{A^3}{\sqrt{N}}$. Consequently, a significant decrease of the sample losses is expected by using MLCs as compared to SLCs covering a comparable surface area. However, this estimate does not account for the increase of the total sample losses due to mutual resistances between the loops. Indeed, the use of N loops in series will add $N \times (N-1)$ mutual resistances, having different amplitude depending on the distance between the considered coupled loops. As the amplitudes of the mutual resistances cannot be estimated for arbitrary MLC geometries, the noise correlation effect was not accounted for in the rough estimation of the gain in sensitivity expected by using MLCs as compared to SLCs since it aims only at illustrating the concept supporting the present investigation.

Estimation of Coil Losses

According to Equations (15) and (16), the ohmic resistance associated to the coil winding of N loops of radius $a = \frac{A}{\sqrt{N}}$ in series is $\sqrt{N}$-times higher than that of an SLC of radius A assuming identical wire radius.

The equivalent series resistance of the capacitors depends on the number of capacitors which is proportional to the total conductor length so as to maintain the same ratio between operating wavelength and uninterrupted conductor length. Since the total length of N small loops of radius $a = \frac{A}{\sqrt{N}}$ in series is $\sqrt{N}$-times longer than that of a loop of radius A, the number of required capacitors for the N small loops is $\sqrt{N}$-times higher than for the single large loop. So, at a first glance, it could be concluded that the resistance of the capacitors for the N small loops will be $\sqrt{N}$-times higher than for the large loop. When using more distributed capacitors the capacitance value has to be increased so as to reach the same resonance frequency. As a general tendency, the higher the capacitance value is, the lower the equivalent series resistance of the capacitors is; therefore, in practice the total resistance of the capacitors in case of N small loops may be increased by a factor lower than $\sqrt{N}$. However, in some cases (depending on the type of capacitors and the operating frequency), the above mentioned tendency does not hold true, i.e., capacitors with lower capacitance value exhibit a lower, or comparable equivalent series resistance to capacitors with higher capacitance value. In conclusion, the total coil resistance, including ohmic and capacitors losses, of N loops of radius $a = \frac{A}{\sqrt{N}}$ in series is increased by a factor of $\sqrt{N}$ as compared to that of a loop with radius A.

Estimation of the Sensitivity Factor

Taking into account the above considerations on the influence of the number of loops on the resistances and the induction coefficient of the coils, one can estimate the global impact of using MLCs on the sensitivity factor. To do so, two cases are distinguished.

At short distances inside the sample, the induction coefficient of the MLC is $\sqrt{N}$-times higher than the one of an SLC of radius A.

In this case, the sensitivity factor achieved by MLCs, expressed as a function of the single loop parameters $\left(\left(\frac{B_1}{I}\right)_A, R_{SA}, R_{CA}\right)$, is:

$$S_{RF(MLC)} = \left(\frac{B_1}{I}\right)_A \sqrt{\frac{N\sqrt{N}}{R_{SA} + R_{CA}N}} \tag{17}$$

At long distances inside the sample, the induction coefficient of the MLC is comparable to the one of the SLC of radius A, and the sensitivity factor achieved by the MLC is:

$$S_{RF(MLC)} = \left(\frac{B_1}{I}\right)_A \sqrt{\frac{\sqrt{N}}{R_{SA} + R_{CA}N}} \tag{18}$$

As expected, the benefit of using an MLC as compared to a large SLC depends on the respective contribution of coil losses and sample induced losses and will be more pronounced when the sample losses are dominant. A rough analysis of the loss dependency on coil radius and operating frequency indicates that sample noise is dominant for large coils or at high frequency. In this case, MLCs can achieve a significant sensitivity improvement. In the other case, when coil losses contribute significantly to the total losses, the benefit of using MLCs is less but a non-negligible improvement may be still expected.

If the sample losses largely dominate over coil losses, the sensitivity factor of the MLC is increased in comparison to the SLC by a factor of $N^{3/4}$ and a factor of $N^{1/4}$ at short and long distances inside the sample, respectively.

As an intermediate case, considering that the coil losses are equal to the sample induced losses for a loop of radius A, i.e., $R_{SA} = R_{CA}$ then at short distances:

$$S_{RF(MLC)} = S_{RF(A)}\sqrt{\frac{N}{N+1}}\sqrt{2} \tag{19}$$

And at long distances:

$$S_{RF(MLC)} = S_{RF(A)}\frac{\sqrt{2}}{\sqrt{N+1}} \tag{20}$$

In the extreme case, when coil losses dominate over sample losses, corresponding to the less favorable case for using MLCs, the sensitivity factor achieved by the MLC as compared to the SLC is increased by a factor of $N^{1/4}$ and decreased by a factor of $N^{1/4}$ at short and long distances inside the sample, respectively.

As a general conclusion of the rough estimate of the expected improvement in sensitivity achieved by the use of MLCs, it appears that an MLC performs better than a large SLC in any case except when the coil noise of the large loop dominates over the sample induced noise for ROIs deep inside the sample.

MATERIALS AND METHODS

Investigated Coils

We illustrate the sensitivity improvement achieved by the MLC principle by investigating two MLCs made of $N = 19$ loops in series. The general scheme and dimensions of the two investigated MLCs are shown in **Figure 1A**. The performance of each MLC is compared to that of an SLC shaving the same outer diameter. The scheme of the SLCs used for comparison is depicted in **Figure 1B**. So as to reduce capacitively coupled sample noise, the winding of the MLCs and the SLCs were segmented by 12 and 8 or 4 (for 3 T or 7 T SLCs) distributed tuning capacitors, respectively. In all cases, it is ensured that the segment length between two capacitors is small as compared to the operating wavelength. Within these constraints, the exact number of capacitors was chosen also according to practical reasons, e.g., reusing existing coil layouts. The two straight lines connected at the bottom of the coils' winding are for connecting the coaxial cable that relates the coil to the scanner interface, where each has a gap for placing balanced matching capacitors. The coil diameters were chosen so that the sample induced losses in the SLC can be assumed large as compared to the internal losses of the coil itself at the operating frequency [1], thus allowing to better evidence the sensitivity improvement achieved by the use of MLCs.

All coils were fabricated using single layer copper of 35 µm thickness deposited on a 0.8 mm thick FR4 substrate. The conductor width was 2 mm for all coils. The coil patterns were produced using standard photolithographic processing.

Evaluation by Simulation

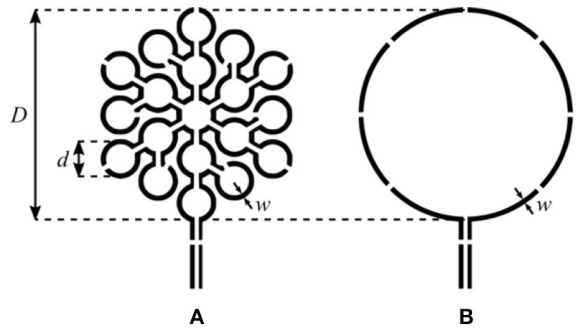
Analytical Estimation of the RF Sensitivity Factor Using the Quasi-Static Approximation

In order to evaluate more precisely the expected RF sensitivity factor of MLCs and validate the initial rough estimate of the sensitivity improvement on which the MLC concept is based, S_{RF} maps were computed for the investigated MLCs and SLCs according to Equation (1) using Python. The induction coefficient was calculated using Equations (2)–(4). For the coil noise, conductor losses including the skin effect were estimated according to Equation (16); it is assumed that a more advanced model for conductor losses including lateral skin effect and proximity effect would only marginally influence the results as the total coil noise is dominated by capacitor and solder joint losses, and the overall noise is clearly sample dominated. Capacitor losses were modeled using the equivalent series resistances extracted from data sheets (CHB series, Exxelia, Paris, France), and solder joint resistances were estimated according to literature data [23], which were extrapolated with a $\sqrt{f}$-dependence for 297.2 MHz. The sample noise of the individual loops was calculated using Equation (12). In order to account for noise correlation between loops in MLCs, R_{Sik} values were estimated using approximated electrical coupling coefficients based on values determined by Roemer for square-shaped loops [6]. Interactions between all coils were accounted for, resulting in a total sample induced resistance being 60 times the sample induced resistance of one isolated small loop. When neglecting the noise correlation effect, as for the rough estimate of the RF sensitivity improvement discussed above, the total sample induced resistance is 19 times the one of an isolated small loop; i.e., noise correlation increases sample losses approximately by a factor of 3 for the investigated MLC design.

Fullwave 3D Electromagnetic Simulation

Due to the complex interactions between electromagnetic fields and the human body, especially at ultra-high frequency i.e., at ultra-high field strength, the quasi-static approximation is no longer valid and Biot-Savart law fails to accurately evaluate the B_1 field of the coil. For this reason, fullwave 3D electromagnetic simulation (EMS) of RF coils has become mandatory to characterize their performances before fabrication. 3D EMS solve Maxwell's equations to obtain the electric and magnetic field distributions inside the sample that can further be used to calculate B_1^+ (transmit) and B_1^- (receive) fields and the specific absorption rate (SAR).

In this study, we performed fullwave 3D EMS based on the finite difference time domain (FDTD) method [29] using a commercial software package (XFdtd 7.8 Remcom, State College, PA, USA). All investigated coils were modeled as perfectly conducting sheet bodies in 3D EMS. A box-shaped phantom positioned 1 mm below the coil was used as load (3 T: 170 × 170 × 150 mm; 7 T: 90 × 90 × 70 mm), with dielectric properties comparable to the phantom liquid used for the experimental evaluation described below (electrical conductivity $\sigma = 0.71$ S/m, relative permittivity $\varepsilon = 63.86$). Grid resolution varied from 0.5 mm for the coil conductors to 9 mm for regions outside the sample. All coil capacitors as well as the matching networks were replaced by 50 Ω voltage sources to enable circuit co-simulation [30, 31] (ADS, Keysight Technologies, USA), which shortens the total simulation time. In co-simulation, 50 Ω ports were replaced by respective lumped elements for tuning and impedance matching. Realistic loss estimations for



3 T coils:
$d = 2$ cm, $D = 12$ cm, $w = 2$ mm
$f_0 = 123.2$ MHz

7 T coils:
$d = 1$ cm, $D = 6$ cm, $w = 2$ mm
$f_0 = 297.2$ MHz

FIGURE 1 | (A) General scheme of the investigated MLCs. Each MLC is composed of 19 loops of diameter *d* connected in series covering an equivalent circular surface of diameter *D*. **(B)** Scheme of the SLCs having a diameter *D* used for comparison. For all coils, the gaps in the coil winding are made for placing the distributed tuning capacitor, and the gaps on the two straight lines at the bottom of the coils' winding are made for placing the balanced matching capacitors.

the coil conductors, inductances, capacitances, and solder joints were modeled as resistances in series with the coil winding, in accordance with coil noise calculation for the analytical estimation. The air core inductor required for matching the 3 T SLC (see below) was assigned a Q of 200. Post-processing of the simulation data was performed in Matlab (Mathworks, Natick, MA, USA) using a dedicated in-house toolbox (SimOpTx, Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Austria) employing the quadratic form power correlation matrix formalism [32, 33]. MLCs and SLCs were compared in terms of transmit efficiency, i.e., B_1^+ per input power, as well as 10 g-averaged SAR.

Experimental Evaluation

Phantom

A canister (∼18 × 13 × 28 cm) filled with saline solution (deionized H_2O doped with 0.8 mL/L Gadolinium solution with a concentration of 279.32 mg/mL of Gadoteric acid, and 4 g/L NaCl resulting in a DC conductivity of $\sigma = 0.65$ S/m) was used as phantom load for bench measurements as well as for MR experiments at 3 T and 7 T. A photograph of the phantom with the 7 T MLC attached to it is shown in **Figure 2**.

FIGURE 2 | Photograph of the phantom with the 7 T MLC attached to it.

Bench Measurements

Tuning and matching of the coils were performed using small multi-layer non-magnetic high-Q ceramic capacitors (CHB series, Exxelia, Paris, France). Besides their MR compatibility, these capacitors were chosen to minimize the total associated ESR. For each coil, series tuning capacitors C_S, a parallel tuning and matching capacitor C_{TM} and two identical series matching capacitors C_M with the values listed in **Table 1** were used. In order to match the 3 T SLC to 50 Ω, an additional parallel inductor L_M had to be incorporated in the matching network of this coil. As the coils were designed to operate in the sample noise dominated regime, this low-Q (∼200) inductor is assumed to have negligible influence on the coil's MR performance.

All coils were connected to RG316 SPC coaxial cables (AXON' CABLE S.A.S. Montmirail, France), which served as connection to the two-port vector network analyzer (E5071C, Agilent, Santa Clara, USA) in bench measurements and as connection to the scanner interface in MR experiments.

Besides tuning and matching, also Q-factors of all investigated coils were measured in loaded and unloaded condition. This enabled us to ensure that sample noise is the dominant noise source as aimed for in this study, and to evaluate the sample noise reduction achieved by MLCs. Q-factors were measured using the single-loop probe method [34], while the coils were not connected to their respective matching networks. The influence of the single-loop probe was considered negligible when the reflection coefficient at its terminal was below −40 dB.

TABLE 1 | Tuning and matching components.

	C_S [pF]	C_{TM} [pF]	C_M [pF]	L_M [nH]
MLC−3 T	33 (10x) 33 + 1.8 (1x)	27	56	–
SLC−3 T	18 + 18	22 + 12	56 + 68	70
MLC−7 T	12 (8x) 15 (3x)	8.2	18	–
SLC−7 T	2.7 + 2.7 (1x) 5.6 (2x)	5.6	39	–

TABLE 2 | Q-factors.

		MLC−3 T	SLC−3 T	MLC−7 T	SLC−7 T
Bench	$Q_{unloaded}$	163	203	176	268
	Q_{loaded}	33	9	30	7.2
	$Q_{unloaded}/Q_{loaded}$	4.9	22.6	5.9	37.2
Analytical	$Q_{unloaded}$	226.1	236.8	199.5	424.4
	Q_{loaded}	40.3	5.9	68.4	11
	$Q_{unloaded}/Q_{loaded}$	5.6	40.1	2.9	38.6
3D sim.	$Q_{unloaded}$	200.3	231.3	193.6	297.2
	Q_{loaded}	33.8	8	25.4	6.3
	$Q_{unloaded}/Q_{loaded}$	5.9	28.9	7.6	46.9

MRI Experiments

Both, 3 T and 7 T MR experiments were carried out on whole-body MRI systems (3T Prisma Fit and Magnetom 7T MRI, Siemens Healthcare, Erlangen, Germany), with the MLCs and SLCs operated in transmit-receive mode. For this purpose, a home built (for 3 T) and a third party (for 7 T; Stark Contrast, Erlangen, Germany) transmit-receive (T/R) switch with integrated low-noise preamplifiers (3 T: 0.5 dB noise figure, 27.0 ± 0.1 dB gain, Hi-Q.A. Inc., Carleton Place, Ontario, Canada; 7 T: 0.5 dB noise figure, 27.2 ± 0.2 dB gain, Siemens Healthcare, Erlangen, Germany) were used. At both field strengths, the same T/R switch and preamplifier were used for MLC and SLC, respectively, to avoid an influence of the interface components on the comparison.

With all investigated coils, flip angle maps were acquired in 13 coronal slices parallel to the coil plane using the saturated Turbo FLASH (satTFL) method [35]. The first slice was positioned directly at the phantom surface as close as possible to the coils; slice thickness was 3 mm and a spacing of 2 mm between consecutive slices was chosen. As the B_1^+ field of surface coils decreases rapidly along the coil axis, different amplitudes of the saturation pulse were chosen for the different slices in order to generate flip angles in the usable range [36]. The following sequence parameters were used at 3 T: repetition time T_R = 12.64 s, echo time T_E = 2.64 ms, 192 × 192 acquisition matrix, 1.5 mm × 1.5 mm in-plane pixel size, 1 average. At 7 T, the following parameters were applied: T_R = 6.93 s, T_E = 2.66 ms, 128 × 128 acquisition matrix, 1.5 × 1.5 mm in-plane pixel size, 4 averages. From measured flip angle distributions, B_1^+ maps normalized to the input power were calculated using an in-house written Matlab script (Mathworks, Natick, MA, USA) taking into account an insertion loss of −2 dB of the coil cables and the T/R-switches, determined on the bench.

High-resolution 3D gradient echo (GRE) sequences were employed to evaluate the imaging performance of the investigated coils. Data from these scans were used to calculate SNR maps in the central sagittal slice (i.e., in the yz-plane with B_0 along the z-direction and the coil parallel to the xy-plane) with the basic ROI method for comparison of MLCs and SLCs. For 3 T measurements, the following sequence parameters were used: T_R = 6.8 ms, T_E = 2.88 ms, 288 × 234 × 224 acquisition matrix, 1 mm isotropic pixel size, flip angle α = 5°, pixel bandwidth BW = 545 Hz/Px, T_{acq} = 5:56 min. The sequence used in 7 T experiments had the following parameters: T_R = 15 ms, T_E = 6.86 ms, 256 × 256 × 128 acquisition matrix, 1 mm isotropic pixel size, α = 8°, BW = 100 Hz/Px, T_{acq} = 4:36 min.

RESULTS

Q-Factors

Q-factors of the investigated MLCs and SLCs in unloaded and loaded condition are summarized in **Table 2** for 3 T and 7 T, respectively. For comparison, also the Q-factors obtained from analytical calculations and fullwave simulations are listed. Further, the ratio of unloaded to loaded Q-values is calculated.

The measured Q-factors reflect well the behavior expected from theoretical considerations described in sections Estimation of sample losses and Estimation of coil losses. The unloaded Q-factor, inversely proportional to the coil noise, is lower for MLC than for SLCs. However, the Q-ratios show that the overall noise of the experiment, i.e., coil noise plus sample noise, is clearly dominated by the sample noise for all investigated configurations, as aimed for in the coil design process. The loaded Q-factors are higher for MLCs than for SLCs by factors of 3.67 and 4.17 at 3 T and at 7 T, respectively, demonstrating the sample noise reduction by employing the MLC principle. These ratios approach the theoretically expected reduction factor for the sample noise, i.e., $\sqrt{19}$ = 4.36, which was estimated under the approximations described above, and which could only be measured for pure sample noise dominance.

Sensitivity Factor and Transmit Efficiency

Maps of the calculated sensitivity factor, as well as the simulated and measured transmit efficiency, i.e., B_1^+ normalized to the input power ($B_1^+/\sqrt{P}$), at different depths inside the sample, are summarized in **Figures 3**, **4** for 3 T and 7 T coils, respectively. The slice locations shown for calculations and simulations were chosen to correspond to the experimental data; for slices located further away from the coil, experimental B_1^+ maps appear too noisy for a visual comparison due to insufficient SNR of the flip angle mapping sequence. The analytical calculation does not aim at providing absolute S_{RF} values but at estimating the expected sensitivity improvement. Therefore, all the sensitivity factors calculated analytically were normalized to the maximum

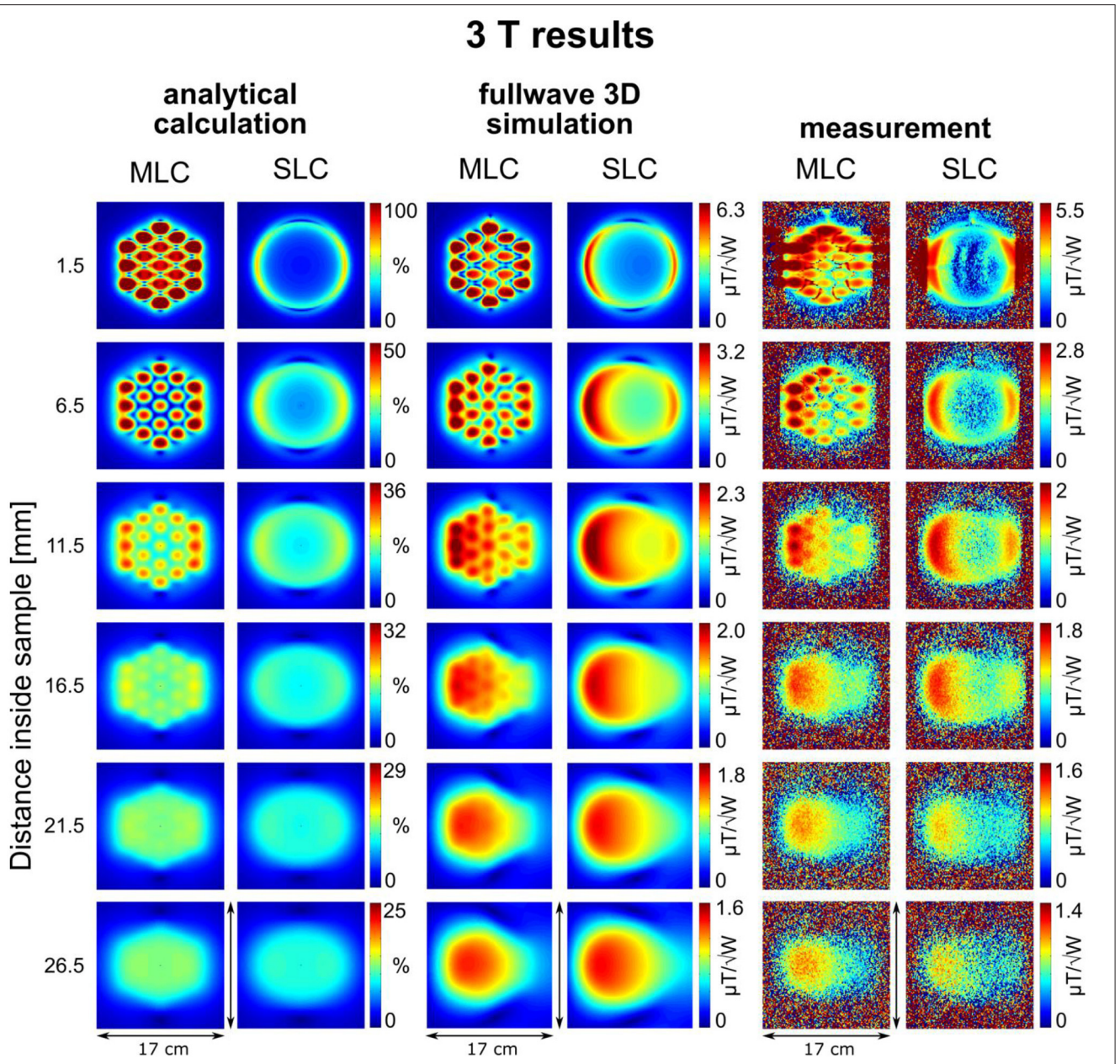


FIGURE 3 | Maps of the calculated sensitivity factor, as well as the simulated and measured transmit efficiency at different depths inside the sample for 3 T coils. Calculated sensitivity factor values are normalized by the maximum sensitivity value obtained with the MLC at 1.5 mm distance inside the sample.

value obtained with the MLCs at the shortest distance inside the sample.

To enable a direct, quantitative comparison of MLCs and SLCs, ratios of MLCs' and SLCs' sensitivity factors and transmit efficiencies were calculated along the central axis of the investigated coils. These results are shown in **Figure 5** for analytical calculations, fullwave 3D simulation and experimental data, for 3 T and 7 T coils, respectively. For experimental data, B_1^+ values in each slice were averaged over 10 × 10 and 5 × 5 pixel ROIs centered on the coil axis, for 3 T and 7 T, respectively, to limit the influence of measurement noise.

An excellent qualitative agreement between fullwave simulations and measurements can be observed for maps as well as central axis profiles, at both, 3 T and 7 T. A maximum increase in transmit efficiency by a factor between 2 and 4 depending on field strength and coil size is obtained with MLCs in comparison to SLCs. A significant gain can be observed for distances up to the radius $d/2$ of the individual loops of the MLCs, i.e., 1 cm for 3 T and 0.5 cm for 7 T. For large distances ($> d$), the performance

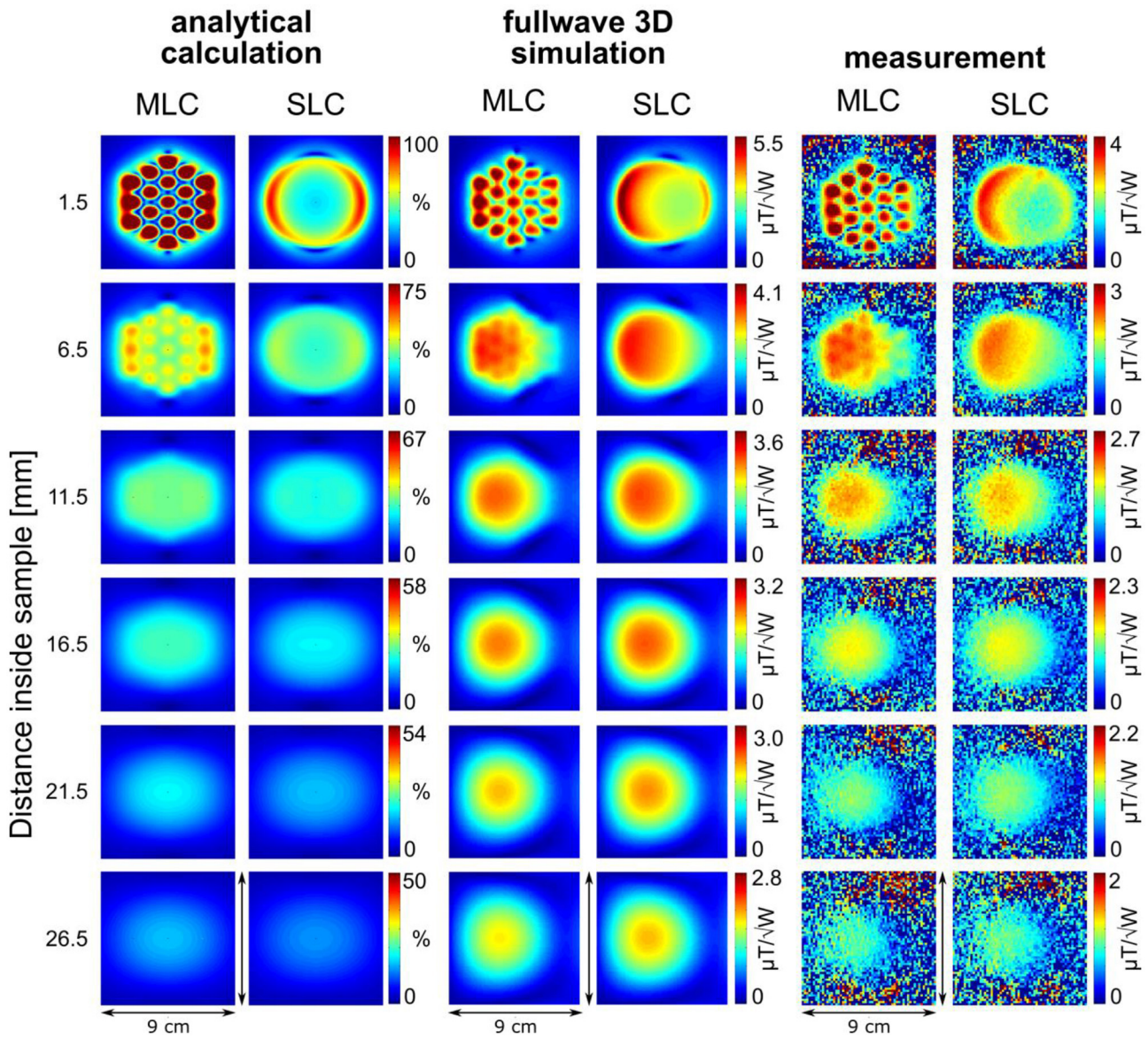


FIGURE 4 | Maps of the calculated sensitivity factor, as well as the simulated and measured transmit efficiency at different depths inside the sample for 7 T coils. Calculated sensitivity factor values are normalized by the maximum sensitivity value obtained with the MLC at 1.5 mm distance inside the sample. Note that the dimensions of the maps are different from 3 T images due to the smaller coil size.

of MLCs and SLCs is comparable. For fullwave simulated data, the SLC marginally outperforms the MLC for distances larger than 3.4 cm at 3 T, and 1.2 cm at 7 T, respectively.

For MLCs, regions of high transmit efficiency directly below the small loops and low transmit efficiency in between loops can be observed in $B_1^+/\sqrt{P}$ maps. The impact of this behavior on the coils' performance in comparison to SLCs was analyzed more closely for 3D fullwave simulation data. The transmit efficiency ratio of MLC and SLC was computed for the whole phantom volume. The central sagittal and transversal slices as well as a coronal slice close to the coil plane are shown in **Figures 6A**, **7A** for 3 T and 7 T, respectively. For the ROI shown by the black dashed line, the fraction of voxels in the ROI with $B_1^+/\sqrt{P}_{MLC} \geq B_1^+/\sqrt{P}_{SLC}$ and the fraction with $B_1^+/\sqrt{P}_{MLC} < B_1^+/\sqrt{P}_{SLC}$ were calculated for each slice up to the distance of break-even along the central coil axis. Also, the mean ratio was calculated for each fraction. Thus, the bar charts in **Figures 6B**, **7B** show which coil (MLC or SLC) performs better in which fraction of voxels in

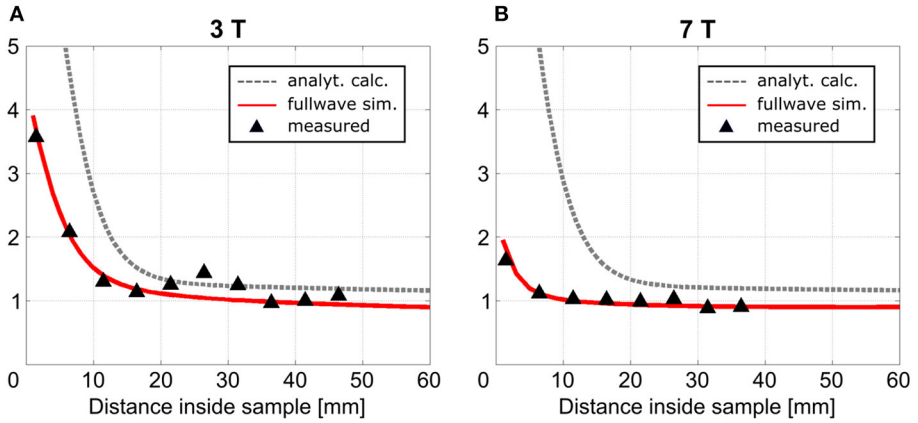


FIGURE 5 | Ratios of RF sensitivity and transmit efficiency between MLCs and SLCs along the central coil axis—**(A)** for 3 T coils and **(B)** for 7 T coils.

each slice. In addition, the color of the bars shows how big the difference in transmit efficiency is; dark colors indicate that the calculated mean ratio in the respective fraction is clearly bigger or smaller than 1; in contrast, light colors indicate a ratio close to 1 (with white corresponding to 1 ± 5%). Further, axis profiles of the ratio through regions of high ("case-high") and low ("case-low") transmit efficiency of MLCs are shown in **Figures 6C**, **7C** for 3 T and 7 T, respectively. It can be observed that the "case-low" profile approaches the value of 1 for smaller distances than the "case-high" profile, which indicates a potential net sensitivity gain with MLCs.

A quantitative comparison of fullwave simulation and measurement results reveals that the values extracted from experimental data are ∼15 and 25% lower than the simulated values for 3 T and 7 T, respectively. There are several potential reasons for this. The most plausible explanation, in our opinion, is a mismatch in sample conductivity between experiment and simulation. In simulation, the sample conductivity was fixed to the value given in the methods section, while the conductivity of the phantom solution was determined with a simple DC probe. However, the conductivity of saline solution generally increases with frequency [37]. Thus, it can be assumed that the conductivity of the phantom is higher in experiments than in simulations, which would result in a stronger dampening of the B_1^+ inside the sample. Another reason for the discrepancy between simulation and measurement could be an underestimation of losses, either in the coil (as it can be seen from the measured and simulated Q-values summarized in **Table 2**) or in the interface components.

For 3 T experiments, the B_1^+ patterns shown in the first slice of **Figure 3** appear smeared; further, a strong edge between regions inside and outside the phantom can be observed in this slice. The reason for this is, that the walls of the phantom canister are very thin, and that, therefore, the shape of the phantom is not perfectly rectangular, but slightly curved. When the large 3 T coils with an outer diameter of 12 cm were attached to the phantom with adhesive tape, the coil PCBs were slightly bent so as to perfectly match the phantom surface. This effect is not observed on the images acquired at 7 T because the overall dimension of the 7 T coils (6 cm) is small enough to produce negligible bending when the coils were attached to the canister.

For analytical calculations, results deviate more from those obtained by the other methods. A stronger decrease of S_{RF} with increasing distance from the coil occurs, the performance of the SLCs is clearly underestimated in comparison to the MLCs, and the left-right asymmetry depicted in fullwave simulation and measurements is not observable. This can be explained by the limitations that apply for the analytical approach. Firstly, the used equations are valid in the quasi-static domain only and, therefore, do not account for propagation effects of the RF EM field that can occur at high frequency, e.g., the proton Larmor frequency at 7 T. This explains the larger deviation observed at 7 T, especially the left-right asymmetry [38]. Secondly, for MLCs, the analytical computation was done considering small loops carrying equal current without accounting for the small conducting lines that connect the individual loops, without exactly calculating mutual resistances, and without accounting for mutual inductances between the loops that tend to reduce the magnetic efficiency; thus, the performance of the MLCs is likely overestimated in analytical calculations.

Nonetheless, the behavior, that MLCs produce a higher B_1^+ field than SLCs in regions close to the coil (especially directly below the small loops) expected from the rough estimation based on theoretical considerations, is well confirmed by analytical calculations, fullwave 3D simulation as well as experimental data. In regions located further away from the coil, B_1^+ strengths of MLCs and SLCs become comparable.

SAR

MLCs and SLCs were also compared in terms of SAR values obtained from fullwave simulation data, in order to assess the

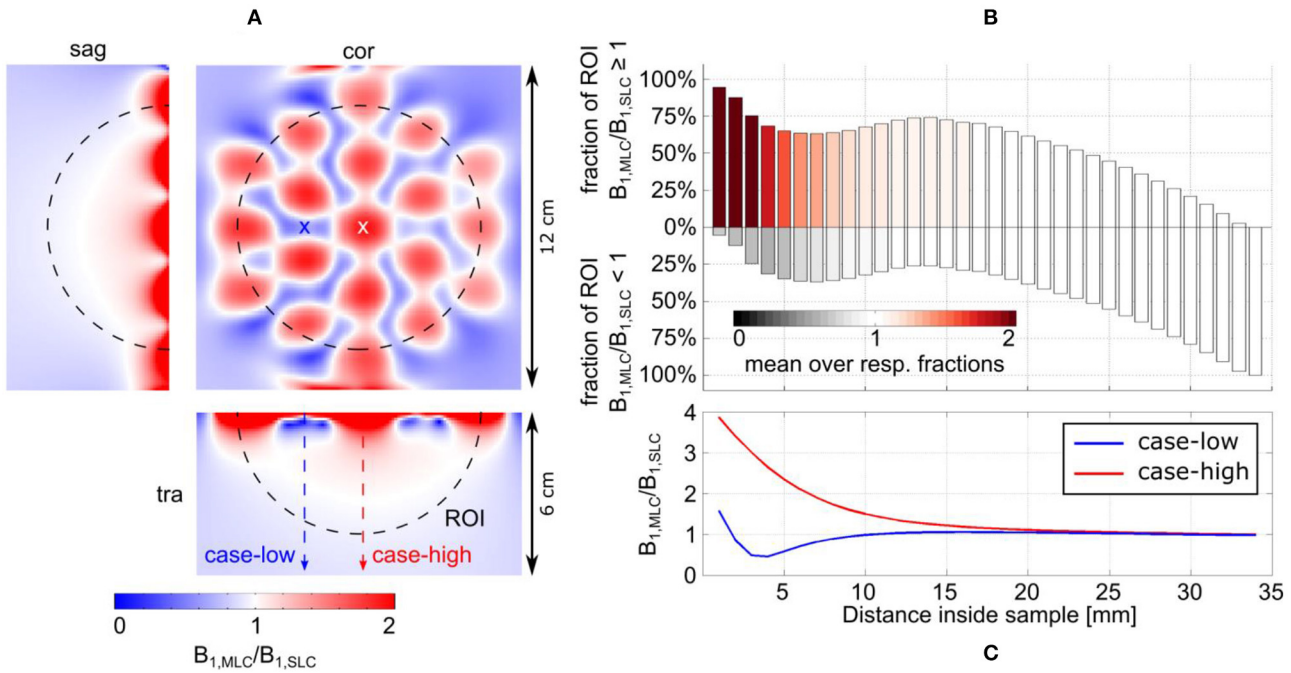


FIGURE 6 | Detailed analysis of the transmit efficiency ratio between MLC and SLC at 3 T. **(A)** Ratio maps in the central sagittal and transversal slice as well as a coronal slice in a depth of 7 mm. **(B)** Analysis of the ratio in the ROI shown by the dashed black line in **(A)**. The position of the bars along the y-axis shows the fractions of the ROI for which either the MLC or the SLC have higher transmit efficiency. The color of the bars indicates how much the ratio in the respective fractions deviates from 1; i.e., white indicates comparable performance (1 ± 5%). **(C)** Line profiles parallel to the coil axis for locations of high (red dashed line and white cross, "case-high") and low (blue dashed line and cross, "case-low") B_1 intensity of the MLC.

usability of MLCs in future *in vivo* studies regarding safety in terms of RF heating. Maximum 10 g-averaged SAR was found to be slightly lower for MLCs than for SLCs, as shown in **Table 3**. This is an interesting finding, especially in the regard, that lower pulse voltages are required with MLCs to generate the same flip angles as SLCs in regions close to the coil, as can be concluded from simulated and measured B_1^+ maps.

SNR in MR Imaging

Figures 8, **9** show SNR maps and corresponding ratio maps obtained from 3D GRE acquisitions for MLCs and SLCs at 3 T and 7 T, respectively. A clear SNR increase can be observed for the MLCs in regions close to the coil, especially directly below the small loops. Further, it can be seen, that the lateral coverage as well as the SNR in deeper lying regions of the phantom (i.e., further away from the coil) are comparable for MLCs and SLCs.

DISCUSSION AND CONCLUSION

Discussion

In this study, we have introduced the MLC principle, and we have investigated the sensitivity improvement that can be achieved in comparison to SLCs with the same overall dimensions. This comparison was done using different methods. Starting from a rough estimation, the expected gain was evaluated more precisely using analytical formulae and, finally, determined using fullwave EM simulations and MR experiments. Results from all employed methods consistently show a strong sensitivity gain with MLCs over SLCs in regions close to the coil (approximately up to the radius of the individual loops in the MLC), especially directly below the small loops, and comparable performance for regions located further away from the coil.

While in this work we have investigated MLCs composed of 19 loops with two different sizes operating at 123.2 MHz and 297.2 MHz, the obtained results are representative for the general sensitivity improvement that can be achieved by using the MLC design. Depending on the operating frequency and the desired FoV, MLCs with different number of loops or with different loop diameter or shape can also be advantageous. The highest benefit of MLCs is observed when sample induced noise dominates. As a general trend, this is the case when using surface coils larger than 3 cm in diameter and operating at static field strengths of 1.5 T and above, i.e., for most of the common settings encountered in clinical MRI applications. It should be noted that the sensitivity improvement achievable with increasing the number of small loops is limited by noise correlation. This limit will be reached when the total sample related resistance will be more increased due to noise correlation than the sample noise reduction achieved by using smaller loops in series ($\sqrt{N}$). As

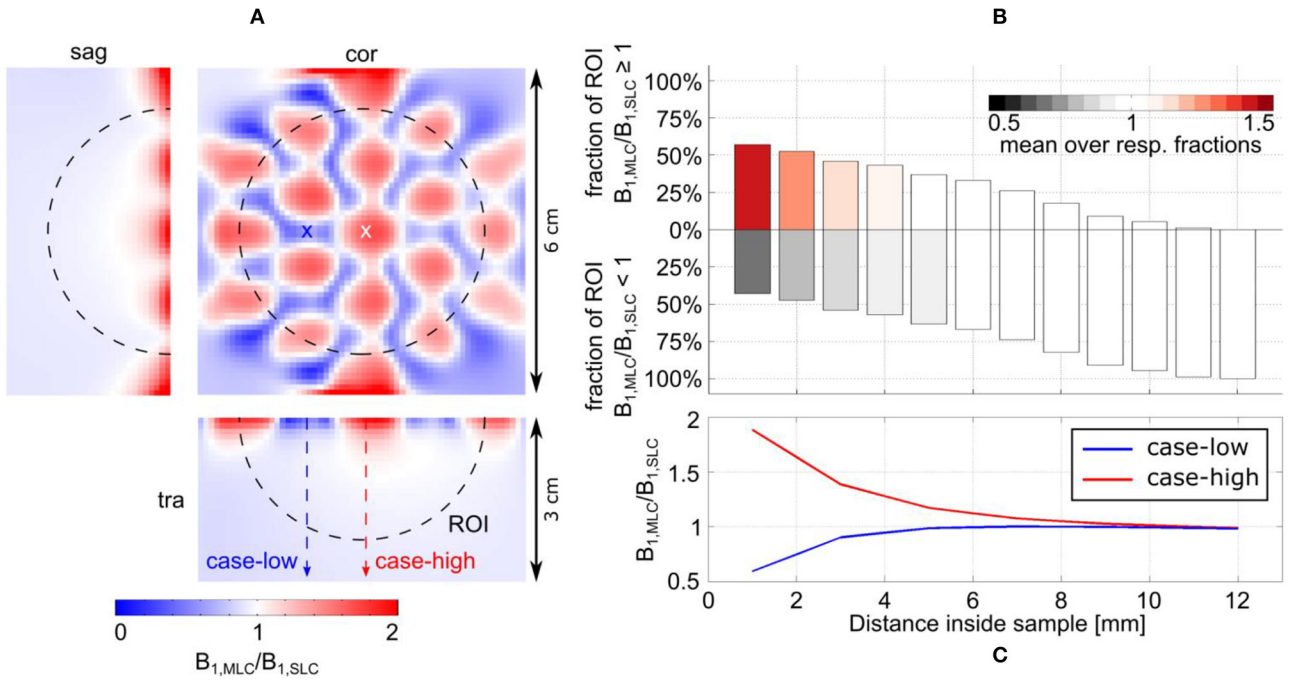


FIGURE 7 | Detailed analysis of the transmit efficiency ratio between MLC and SLC at 7 T. **(A)** Ratio maps in the central sagittal and transversal slice as well as a coronal slice in a depth of 2 mm. **(B)** Analysis of the ratio in the ROI shown by the dashed black line in **(A)**. The position of the bars along the y-axis shows the fractions of the ROI for which either the MLC or the SLC have higher transmit efficiency. The color of the bars indicates how much the ratio in the respective fractions deviates from 1; i.e., white indicates comparable performance (1 ± 5%). **(C)** Line profiles parallel to the coil axis for locations of high (red dashed line and white cross, "case-high") and low (blue dashed line and cross, "case-low") B_1 intensity of the MLC.

a perspective, for configurations where the use of small loops in series is advantageous in the presence of noise correlation, even higher SNR improvement could be obtained by minimizing noise correlation; for instance, Algarin et al. [39] have recently demonstrated a significant reduction of the electrical coupling coefficient using a metamaterial surface.

Results presented here were obtained using MLCs fabricated from copper clad laminated FR4 substrate, but the MLC principle shows no particular restriction regarding the technology used for coil fabrication and can therefore be applied as well to produce MLCs made of flexible substrates or, more standardly, from wound copper wire.

As compared to the rough estimate of the sensitivity improvement presented in section Rough estimate of the sensitivity improvement expected with MLCs, the RF sensitivity factor computation based on analytical formulae provides a more accurate evaluation and allows for a rapid computation of 3D sensitivity maps that are informative and useful for the MLC design optimization step. However, some limitations apply for this approach as described above; therefore, the use of more advanced simulation methods and the final experimental evaluation are indispensable for RF coil development and evaluation at high and ultra-high field strength.

As it can be observed in sensitivity maps, at very short distances inside the sample, i.e., comparable to the diameters of the individual loops of the MLC, signal loss occurs between adjacent loops of the MLC because of the reverse direction of the B_1^+ field created in this region and because of MR-inefficient B_1 components parallel to B_0. While this phenomenon vanishes further away inside the sample, it might be problematic for MR applications targeting surface ROIs such as skin imaging. The simplest solution in this case would be to offset the coil slightly from the sample in a way to still benefit from the (smaller) SNR gain, but strongly reduce the inhomogeneity. To remedy this issue without offsetting the coil, more complex MLC designs will be investigated in future work so as to achieve current patterns that generate a more uniform B_1 distribution close to the coil. For instance, this could be done by adding smaller loops in series to the initial ones in the regions where signal loss occurs. This approach is possible as the MLC principle brings additional degrees of freedom for coil design as compared to SLCs, which potentially enables the realization of specific coil patterns with

TABLE 3 | Maximum 10 g-averaged SAR values.

	3 T	7 T
MLC	2.0461	5.4518
SLC	2.2654	6.5142
rel. difference	−9.7 %	−16.3 %

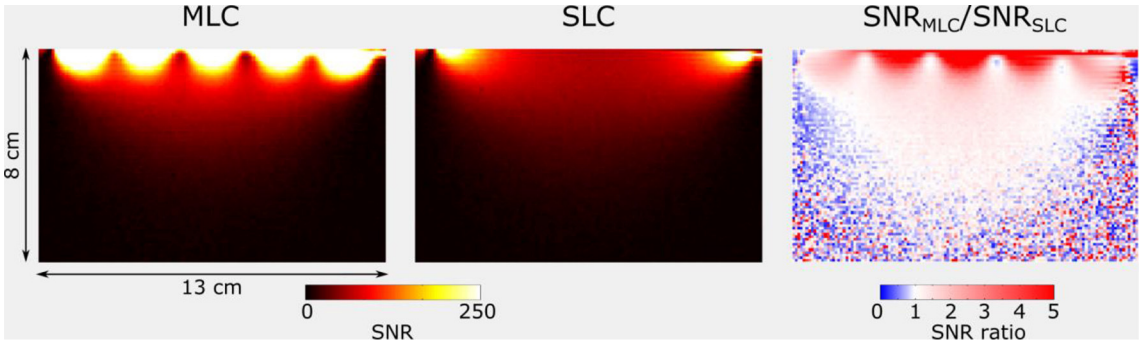


FIGURE 8 | SNR and ratio maps obtained for the central sagittal slice of 3D GRE acquisitions at 3 T for MLC and SLC, respectively. Maps were cropped, so as to remove noise-only regions.

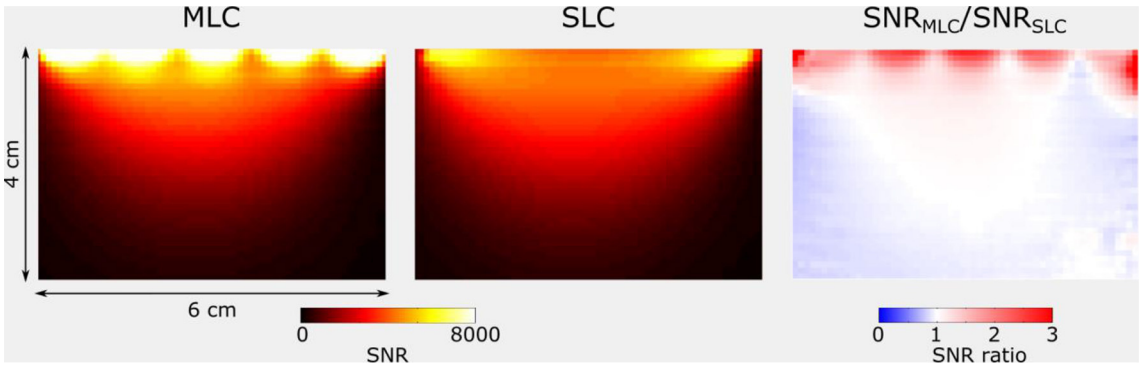


FIGURE 9 | SNR and ratio maps obtained for the central sagittal slice of 3D GRE acquisitions at 7 T for MLC and SLC, respectively. Maps were cropped, so as to remove noise-only regions.

respect to the sample shape, but also the optimization of the homogeneity of the detection sensitivity in a target region inside the sample.

Several MRI applications may well benefit from MLCs, which will be primarily used as single coils for applications requiring high sensitivity over a FoV that is large compared to the target depth, e.g., skin imaging [40], or *ex vivo* imaging of brain slices [41]. In addition, MLCs can also be employed as building block of an array when an even larger FoV has to be covered. The detailed investigation of MLC arrays is subject to future studies.

As compared to arrays of SLCs, the MLC principle brings simplicity for both, the design and the fabrication, while aiming at achieving a comparable performance in terms of sensitivity. This renders possible a significant cost reduction that may strongly inure to the benefit of developing countries where very high-priced parallel acquisition MRI systems appear unaffordable. On the other hand, using a single MLC instead of an array of SLCs is not compatible with parallel imaging approaches for accelerated image acquisition and does not allow for SNR optimization using a weighted signal combination of the individual channels. However, for those applications expected to benefit most from the MLC principle, targeting very high resolution in shallow depth over a large FoV, sensitivity is typically more important than acquisition speed.

Conclusion

In this paper, the proof of concept of a novel RF coil design, the multi-loop coil design, has been established. The MLC concept exploits the intrinsically high sensitivity of small surface coils that achieve strong magnetic coupling to the sample while reducing the sample induced noise, together with achieving a large FoV by associating multiple small loops in series. It allows for significant sensitivity improvement when the sample induced noise dominates over the internal coil noise, relevant for most clinically applied surface coils ($>$3 cm diameter, $\geq$ 1.5 T).

As a general tendency, close to the coil plane, the MLC design potentially achieves higher RF sensitivity as compared to a single loop coil having the same lateral size. Maximum gains in transmit efficiency by a factor between 2 and 4 were obtained experimentally depending on the field strength and coil size. At long distance from the coil, as the induction coefficients of the individual loops of the MLC are summed

up, the achieved sensitivity is comparable to that of the equivalent SLC.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

J-CG, EL, and RF-K planned the study. J-CG and MP-Q performed the analytical calculation. SH and RF-K performed the fullwave simulations. RF-K acquired the experimental data and performed the data analysis. All authors contributed to writing and proof-reading the article.

FUNDING

This work was funded by the Austrian/French OeAD WTZ grant FR 03/2018 and the Austrian Science Fund grant FWF P28059.

ACKNOWLEDGMENTS

The authors thank Ms. Anika Franta and Ms. Hannah Goetz for contributing to this work.

REFERENCES

1. Darrasse L, Ginefri JC. Perspectives with cryogenic RF probes in biomedical MRI. *Biochimie.* (2003) **85**:915–37. doi: 10.1016/j.biochi.2003.09.016
2. Redpath TW. Signal-to-noise ratio in MRI. *Br J Radiol.* (1998) **71**:704–7. doi: 10.1259/bjr.71.847.9771379
3. Kneeland JB, Hyde JS. High-resolution MR imaging with local coils. *Radiology.* (1989) **171**:1–7. doi: 10.1148/radiology.171.1.2648466
4. Ackerman JJH, Grove TH, Wong GG, Gadian DG, Radda GK. Mapping of metabolites in whole animals by 31P NMR using surface coils. *Nature.* (1980) **283**:167–70. doi: 10.1038/283167a0
5. Bittoun J, Saint-Jalmes H, Querleux BG, Darrasse L, Jolivet O, Peretti II, et al. *In vivo* high-resolution MR imaging of the skin in a whole-body system at 1.5 T. *Radiology.* (1990) **176**:457–60. doi: 10.1148/radiology.176.2.2367660
6. Roemer PB, Edelstein WA, Hayes CE, Souza SP, Mueller OM. The NMR phased array. *Magn Reson Med.* (1990) **16**:192–225. doi: 10.1002/mrm.1910160203
7. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Magn Reson Med.* (1999) **42**:952–62. doi: 10.1002/(SICI)1522-2594(199911)42:5<952::AID-MRM16>3.0.CO;2-S
8. Sodickson DK, Manning WJ. Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. *Magn Reson Med.* (1997) **38**:591–603. doi: 10.1002/mrm.1910380414
9. Heidemann RM, Ozsarlak O, Parizel PM, Michiels J, Kiefer B, Jellus V, et al. A brief review of parallel magnetic resonance imaging. *Eur Radiol.* (2003) **13**:2323–37. doi: 10.1007/s00330-003-1992-7
10. Ohliger MA, Sodickson DK. An introduction to coil array design for parallel MRI. *NMR Biomed.* (2006) **19**:300–15. doi: 10.1002/nbm.1046
11. Dona Lemus OM, Konyer NB, Noseworthy MD. Micro-strip surface coils using fractal geometry for 129Xe lung imaging applications. In: *Proceedings of the International Society for Magnetic Resonance in Medicine*. Paris (2018). p. 1713.
12. Mansfield P. The petal resonator: a new approach to surface coil design for NMR imaging and spectroscopy. *J Phys D Appl Phys.* (1988) **21**:1643–4. doi: 10.1088/0022-3727/21/11/015
13. Rodriguez AO, Hidalgo SS, Rojas R, Barrios FA. Experimental development of a petal resonator surface coil. *Magn Reson Imaging.* (2005) **23**:1027–33. doi: 10.1016/j.mri.2005.09.001
14. Robitaille PM, Berliner L. *Ultra High Field Magnetic Resonance Imaging*. Boston, MA: Springer (2006). doi: 10.1007/978-0-387-49648-1
15. Simpson J, Lane J, Immer C, Youngquist R. *Simple Analytic Expressions for the Magnetic Field of a Circular Current Loop*. (2001). Available online at: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20010038494.pdf (accessed September 10, 2019).
16. Hoult DI, Lauterbur PC. The sensitivity of the zeugmatographic experiment involving human samples. *J Magn Reson.* (1979) **34**:425–33. doi: 10.1016/0022-2364(79)90019-2
17. Decorps M, Blondet P, Reutenauer H, Albrand JP, Remy C. An inductively coupled, series-tuned NMR probe. *J Magn Reson.* (1985) **65**:100–9. doi: 10.1016/0022-2364(85)90378-6
18. Guéron M, Leroy JL. NMR of water protons. The detection of their nuclear-spin noise, and a simple determination of absolute probe sensitivity based on radiation damping. *J Magn Reson.* (1989) **85**:209–15. doi: 10.1016/0022-2364(89)90338-7
19. Kraichman M. Impedance of a circular loop in an infinite conducting medium. *J Res Nat Bur Stand D Radio Propag.* (1962) **66D**:499–503. doi: 10.6028/jres.066D.050
20. Suits BH, Garroway AN, Miller JB. Surface and gradiometer coils near a conducting body: the lift-off effect. *J Magn Reson.* (1998) **135**:373–9. doi: 10.1006/jmre.1998.1608
21. Serfaty S, Darrasse L, Kan S. The pinpoint NMR coil. In: *Proceedings of the SMR Second Annual Meeting*. San Francisco, CA (1994). p. 219.
22. Werner DH. An exact integration procedure for vector potentials of thin circular loop antennas. *IEEE Trans Antennas Propag.* (1996) **44**:157–65. doi: 10.1109/8.481642
23. Kumar A, Edelstein WA, Bottomley PA. Noise figure limits for circular loop MR coils. *Magn Reson Med.* (2009) **61**:1201–9. doi: 10.1002/mrm.21948
24. Giovannetti G, Tiberi G. Skin effect estimation in radiofrequency coils for nuclear magnetic resonance applications. *Appl Magn Reson.* (2016) **47**:601–12. doi: 10.1007/s00723-016-0780-x
25. Giovannetti G, Hartwig V, Landini L, Santarelli MF. Classical and lateral skin effect contributions estimation in strip MR coils. *Concepts Magn Reson B Magn Reson Eng.* (2012) **41B**:57–61. doi: 10.1002/cmr.b.21210
26. Belevitch V. The lateral skin effect in a flat conductor. *Philips tech Rev.* (1971) **32**:221–31.
27. Terman FE. *Radio Engineers' Handbook*. 1st Ed. New York, NY: McGraw-Hill Book Company, Inc. (1943).
28. Gonord P, Kan S, Leroy-Willig A. Parallel-plate split-conductor surface coil: analysis and design. *Magn Reson Med.* (1988) **6**:353–8. doi: 10.1002/mrm.1910060313
29. Yee KS. Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Trans Antennas Propag.* (1966) **14**:302–7. doi: 10.1109/TAP.1966.1138693
30. Kozlov M, Turner R. Fast MRI coil analysis based on 3-D electromagnetic and RF circuit co-simulation. *J Magn Reson.* (2009) **200**:147–52. doi: 10.1016/j.jmr.2009.06.005
31. Lemdiasov RA, Obi AA, Ludwig R. A numerical postprocessing procedure for analyzing radio frequency MRI coils. *Concepts Magn Reson A Magn Reson Eng.* (2011) **38A**:133–47. doi: 10.1002/cmr.a.20217
32. Graesslin I, Homann H, Biederer S, Börnert P, Nehrke K, Vernickel P, et al. A specific absorption rate prediction concept for parallel transmission MR. *Magn Reson Med.* (2012) **68**:1664–74. doi: 10.1002/mrm.24138
33. Kuehne A, Goluch S, Waxmann P, Seifert F, Ittermann B, Moser E, et al. Power balance and loss mechanism analysis in RF transmit coil arrays. *Magn Reson Med.* (2015) **74**:1165–76. doi: 10.1002/mrm.25493

34. Ginefri JC, Durand E, Darrasse L. Quick measurement of nuclear magnetic resonance coil sensitivity with a single-loop probe. *Rev Sci Instrum.* (1999) **70**:4730–1. doi: 10.1063/1.1150142
35. Chung S, Kim D, Breton E, Axel L. Rapid B1+ mapping using a preconditioning RF pulse with TurboFLASH readout. *Magn Reson Med.* (2010) **64**:439–46. doi: 10.1002/mrm.22423
36. Pohmann R, Scheffler K. A theoretical and experimental comparison of different techniques for B1 mapping at very high fields. *NMR Biomed.* (2013) **26**:265–75. doi: 10.1002/nbm.2844
37. Giovannetti G, Frijia F, Menichetti L, Hartwig V, Viti V, Landini L. An efficient method for electrical conductivity measurement in the RF range. *Concepts Magn Reson B Magn Reson Eng.* (2010) **37B**:160–6. doi: 10.1002/cmr.b.20165
38. Collins CM, Wang Z. Calculation of radiofrequency electromagnetic fields and their effects in MRI of human subjects. *Magn Reson Med.* (2011) **65**:1470–82. doi: 10.1002/mrm.22845
39. Algarin JM, Breuer F, Behr VC, Freire MJ. Analysis of the noise correlation in MRI coil arrays loaded with metamaterial magnetoinductive lenses. *IEEE Trans Med Imaging.* (2015) **34**:1148–54. doi: 10.1109/TMI.2014.2377792
40. Laistler E, Loewe R, Moser E. Magnetic resonance microimaging of human skin vasculature *in vivo* at 3 Tesla. *Magn Reson Med.* (2011) **65**:1718–23. doi: 10.1002/mrm.22743
41. Gruber B, Keil B, Witzel T, Nummenmaa A, Wald LL. A 60-channel *ex-vivo* brain-slice coil array for 3T imaging. In *Proceedings of the International Society for Magnetic Resonance in Medicine*. Milan (2014) p. 4885.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

PERSPECTIVE
published: 21 February 2020
doi: 10.3389/fphy.2020.00011

Perspectives in Wireless Radio Frequency Coil Development for Magnetic Resonance Imaging

*Lena Nohava[1,2], Jean-Christophe Ginefri[1], Georges Willoquet[1], Elmar Laistler[2] and Roberta Frass-Kriegl[2]**

[1] Université Paris-Saclay, CEA, CNRS, Inserm, BioMaps, Orsay, France, [2] Division MR Physics, Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Vienna, Austria

This paper addresses the scientific and technological challenges related to the development of wireless radio frequency (RF) coils for magnetic resonance imaging (MRI) based on published literature together with the authors' interpretation and further considerations. Key requirements and possible strategies for the wireless implementation of three important subsystems, namely the MR receive signal chain, control signaling, and on-coil power supply, are presented and discussed. For RF signals of modern MRI setups (e.g., 3 T, 64 RF receive channels), with on-coil digitization and advanced methods for dynamic range (DR $\geq$ 16-bit) and data rate compression, still data rates $>$ 500 Mbps will be required. For wireless high-speed MR data transmission, 60 GHz WiGig and optical wireless communication appear to be suitable strategies; however, on-coil functionality during MRI scans remains to be verified. Besides RF signals, control signals for on-coil components, e.g., active detuning, synchronization to the MR system, and B_0 shimming, have to be managed. Wireless power supply becomes an important issue, especially with a large amount of additional on-coil components. Wireless power transfer systems ($>$10 W) seem to be an attractive solution compared to bulky MR-compatible batteries and energy harvesting with low power output. In our opinion, completely wireless RF coils will ultimately become feasible in the future by combining efficient available strategies from recent scientific advances and novel research. Besides ongoing improvement of all three subsystems, innovations are specifically required regarding wireless technologies, MR compatibility, and wireless power supply.

Keywords: magnetic resonance imaging, radio frequency coil, signal transmission, wireless technologies, wireless power

OPEN ACCESS

Edited by:
Wouter van Elmpt, Maastricht University, Netherlands

Reviewed by:
Simone Angela S. Winkler, Weill Cornell Medicine, Cornell University, United States
Riccardo Lattanzi, Langone Medical Center, New York University, United States

***Correspondence:**
Roberta Frass-Kriegl
roberta.frass@meduniwien.ac.at

Specialty section:
This article was submitted to Medical Physics and Imaging, a section of the journal Frontiers in Physics

Received: *01 October 2019*
Accepted: *09 January 2020*
Published: *21 February 2020*

Citation:
Nohava L, Ginefri J-C, Willoquet G, Laistler E and Frass-Kriegl R (2020) Perspectives in Wireless Radio Frequency Coil Development for Magnetic Resonance Imaging. Front. Phys. 8:11. doi: 10.3389/fphy.2020.00011

INTRODUCTION

Magnetic resonance imaging (MRI) has become one of the major tools in non-invasive medical diagnostics, providing a multitude of quantitative and functional information with ever-increasing performance. The constant search for improved sensitivity and specificity in MR examinations has coined the trend toward MR scanners with higher static magnetic field strength (B_0) [1, 2] and radio frequency (RF) coil arrays with larger numbers of individual receive elements [3]. Today's

high-end clinical MR scanners have a static magnetic field strength of 3 T (together with first clinical 7 T systems being installed currently) and feature up to 64 receive channels (128 or more in some research units), allowing for shorter examination times using parallel imaging [4, 5]. Typically, the excitation of the nuclear spins is done with a large high-power RF transmit coil—the system body coil—included in the scanner bore, while signal detection is performed with a local receive-only coil array, followed by on-coil preamplification and digitization in either the MR room or the technical cabinet, or rarely, on-coil. Coaxial cables are commonly used to transfer the received RF signal to the image reconstruction unit outside the MR scanner room and to power active electronic devices, such as preamplifiers, typically using DC current running on the coaxial cable's shield, which requires a bias-tee arrangement usually already integrated in commercial scanner hardware and thus avoids supplementary power cables. In addition, single wires carrying DC control signals are routed together with the coaxial cables, e.g., to bias PIN diodes as part of an on-coil switching circuitry. With increasing cabling complexity of modern high field scanners equipped with high-density and/or mechanically flexible receive arrays, the use of a large number of coaxial and wire cables gives rise to several challenges.

One main concern with cabling is the increased patient risk due to local heating phenomena associated to currents induced on the cable shields during RF transmission and fast switching of magnetic field gradients [6–8]. Secondly, as each receive element requires its own set of coaxial cable and wires, adjacent routing of cables may lead to cross talk and increase coupling between receive elements, causing a significant reduction of RF detection sensitivity. Since the coaxial cables are routed within the system body coil, a partial loss of transmit power may also occur, as some of the RF power is dissipated in the coil's cabling rather than in the target patient tissue. Baluns and RF traps [9, 10], conventionally used to reduce the abovementioned electromagnetic issues, make the receive coil heavy, bulky, and potentially intimidating and ill-fitting for patients. Moreover, handling of the coil becomes cumbersome and delicate in a way that the coil installation can occupy a significant fraction of the total exam time. This is of particular concern for applications requiring very long coaxial cables, such as abdominal MRI.

Consequently, the use of coaxial cables is one of the bottlenecks that have to be overcome to develop the next generation of coil arrays with improved sensitivity and less patient risk in high field MRI. Several approaches were proposed for the replacement of coaxial receive cables in MR experiments by optical fibers for analog [11–17] or digital [18–24] MR signal transmission. While the use of optical fibers avoids safety issues and reduces signal interferences, the positioning and handling of the receive coils are still limited by the length, placement, and maximum curvature of the optical fibers.

Fully wireless RF coils could lead to a safer, more cost- and time-efficient receive system for MRI and ultimately enable lightweight, flexible, or even "wearable" coil arrays (e.g., [23–26]), improving patient comfort and supporting the evolution of on-coil sensor integration.

Challenges in the development of wireless RF coils can especially be related to the harsh MR environment as all envisioned devices must be designed to be MR compatible, i.e., not ferro- or strongly para-magnetic. Additionally, all parts must function robustly in the strong static B_0 field and handle coil vibrations, patient movement, bore reflections, and most importantly, gradient and RF fields present during MRI. To this end, some sensitive parts can be covered by Faraday cages. Possible current induction on the devices should be avoided with regard to patient safety, and added on-coil devices, e.g., digitization units or wireless transceivers, must appear transparent during imaging. Also, it is desirable to preserve high linearity and a low system noise figure ($<$1 dB [27]) even with the inclusion of wireless technologies. Especially for flexible arrays, a reduction of the total amount, size, and weight of on-coil components is crucial.

In this work, we focus on the realizability of completely wireless MR receive arrays by addressing and interrelating different aspects of the MR receive system. The aim is to outline feasible and efficient approaches toward wireless communication in MRI and prospect digital wireless RF devices, highlighting the most promising strategies as well as associated benefits and challenges.

WIRELESS APPROACHES FOR DIFFERENT PARTS OF THE MR RECEIVE SYSTEM

Three subsystems that have to undergo significant changes for wireless MRI were identified: the MR receive signal chain, control signaling, and on-coil power supply. Their functional blocks and respective possible physical location are depicted in **Figure 1A**. Different wireless transceiver positioning variants, estimated transmission distances, and angles are sketched in **Figure 1B**.

In **Figure 2**, the state of the art in wireless RF coil development, listing existing technologies or strategies for each respective subsystem, corresponding to sections "MR Receive Signal Chain", "Control Signaling", and "On-Coil Power Supply" in the manuscript, is summarized. Specific requirements that need to be met for each of the functional blocks are included, and benefits of current technology as well as current limitations or challenges encountered in their development are listed. The following general requirements apply to all of the mentioned subsystems and corresponding components: MR compatibility (no impact on MRI or component functioning), patient safety (no heating), linearity, low noise figure, low power consumption, low number of additional components, miniature component size, and minimum weight. For all wireless paths, a reliable, ideally lossless, spatial data transmission (≈10–100 cm, see **Figure 1B**) is required.

MR Receive Signal Chain

The features of the MR signal directly impact signal conditioning, which comprises (pre)amplification, digitization, analog and/or digital data compression, and filtering. The MR signal is characterized by high signal frequency (the Larmor frequency),

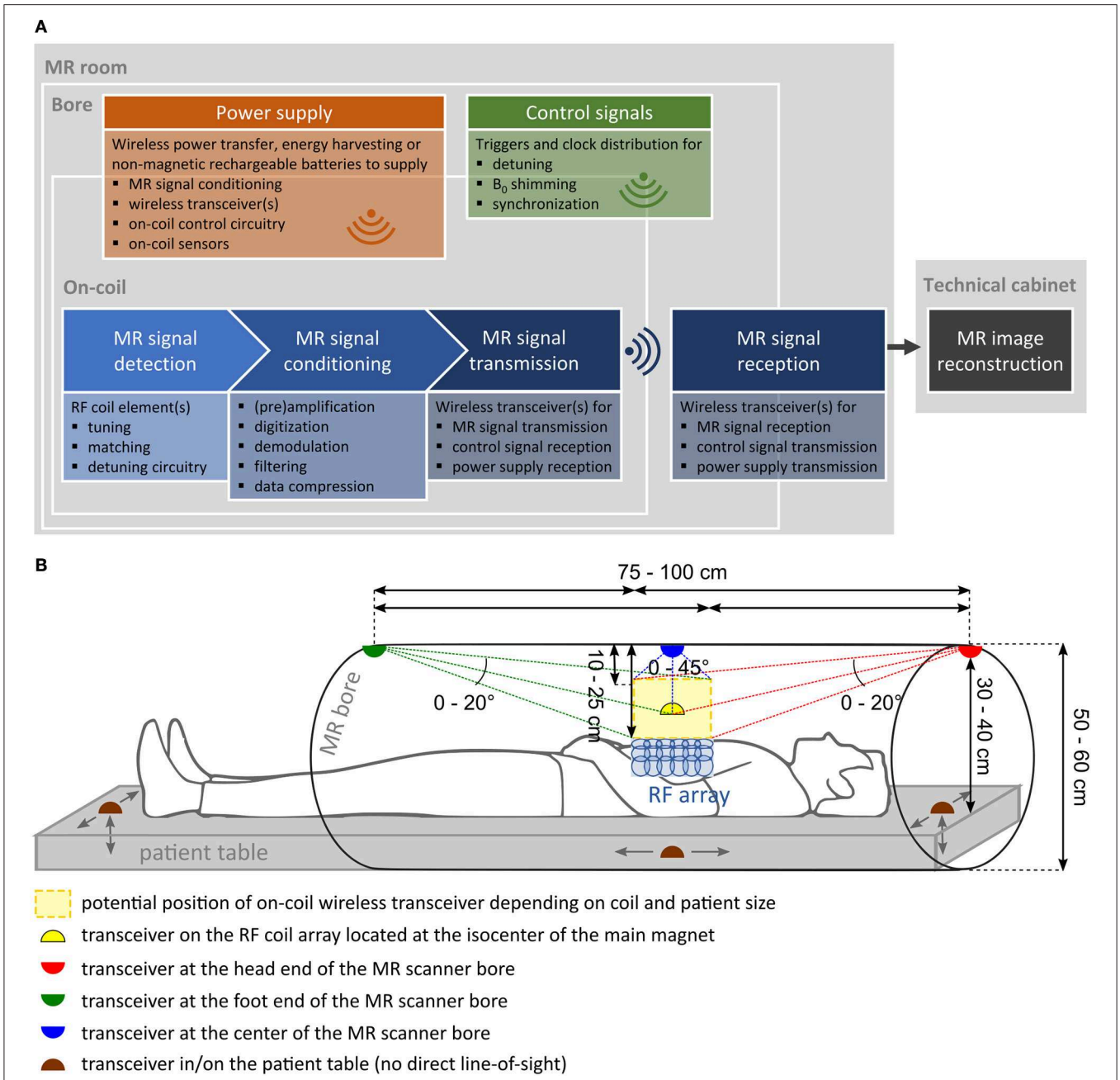


FIGURE 1 | (A) Functional block diagram of a wireless MR receive chain consisting of three main subsystems: wireless MR signals (blue), control signals (green), and power supply (orange). **(B)** Side view of wireless transceiver positioning variants including transmission distance and angle estimations.

depending on the investigated nucleus and B_0 field strength, typically in the order of 50–300 MHz. Further, the DR easily reaches ∼90 dB [28]. In extreme cases, especially for high-resolution 3D acquisitions at high B_0 fields, the DR can attain up to ∼120 dB [29, 30]. To enable proper signal conditioning for various imaging scenarios (frequency, DR, number of receive coil elements, etc.), necessary adaptations for a wireless receive chain imply the relocation of many components inside the MR bore or directly on-coil, e.g., adjustable gain amplifiers, analog-to-digital converters (ADCs), or mixers.

Signal Digitization

The choice of suitable digitization components is a critical task, as there is always a trade-off between achievable conversion rates, bit resolution, power dissipation, cost, and scalability to multi-channel systems. In general, on-coil digitization is advantageous, as it improves signal and phase stability, yielding better image quality, and offers easier scalability to multi-channel systems [18, 19, 31]. For component selection, the main challenges are related to the MR signal properties. Concerning the DR, ADCs should provide high bit resolutions (≥DR in decibels divided by

function	strategies	specific requirements	main benefits (+) and limitations/challenges (-)	Ref.
Input	*preamplified analog filtered MR signal*			
MR signal conditioning				
analog down-conversion	baseband (BB)/ intermediate frequency (IF)	• mixing with local oscillator (LO) signal on-coil • LO sync. to MR system clock	+ lower ADC sampling rate requirement – quadrature (I/Q) mixing necessary for BB conversion – free running LO impaired by gradient inductions	21, 23, 24, 31, 35
digitization	sampling of demodulated MR signal (BB/IF)	• ADC sampling rate $f_s \geq 2*f_{max}$ (except undersampling) • high ADC bit resolution (#b ≥ DR/6.02) • ADC clock synchronization to MR system clock	+ low resulting data rate + easy reconfiguration to other B_0 field strengths by changing LO frequency + availability of low-cost low-speed multi-channel ADCs with high DR + low power consumption + on-coil MR compatibility of ADCs in integrated circuit tested – complexity of circuit design – high amount of on-coil components (using discrete components)	21, 23, 24, 31, 35
	direct (under)sampling without analog down-conversion		+ on-coil MR compatibility of custom high-speed ADCs with high DR tested + low amount of additional on-coil components (no analog mixing) – high resulting data rate – restriction to application at specific B_0 field strength – questionable MR compatibility of many commercially available ADCs – high cost and restricted availability of high-speed ADCs with high DR – high power consumption	18-20, 22, 34, 36-39
digital data compression	(dynamic) demodulation/ filtering/decimation/ DR compression/ bit-depth reduction/ coil compression	• integration of dedicated digital signal processing components in-bore/on-coil • synchronization to MR system clock	+ reduced data for wireless transmission with early on-coil data compression + reduced data storage requirements + easier scalability to systems with many Rx channels/additional sensors – use of hardware on-coil still has to be demonstrated – increase in the number/size of on-coil components (e.g. FPGAs) – increase in required on-coil power	22, 41, 42
Output	*digital (compressed) MR signal*			
wireless MR signals				
MR signal transmission	Wi-Fi (2.4 – 60 GHz carriers)	• achievable data rate ≥ input signal data rate (→ sets requirements for signal conditioning)	+ data rates < 665 Mbps over < 3 m feasible out-of-bore (Wi-Gig dongles) + data rates < 500 Mbps over < 0.65 m feasible in-bore with B_0 only (60 GHz) + low power solutions available (60 GHz) – trade-off between achievable data rate, device size and spatial range – questionable full MR compatibility of wireless transceivers	43, 44, 51
	optical wireless communication (THz carriers)		+ small and low power components + high achievable data rate, e.g. > 3Gbps Li-Fi – MR compatibility of suitable components to be tested – direct line-of-sight often required	56
Output	*wireless MR signal*			
wireless control signals				
active detuning	wireless trigger to switch diodes or FETs	• synchronization to the MR scanner's Tx/Rx state	+ low power FETs available + MR compatibility tested	59, 60
synchron-ization	physical clock transmission	• phase synchronization of multiple receive channels' and on-coil electronics with the MR system	+ only hardware modifications necessary – additional wireless back-channel from MR system to coil required – additional on-coil clocking components required (e.g. PLL)	34, 61
	software synchronization		– often software and hardware modifications necessary – free running on-coil oscillator impaired by gradient inductions – additional wireless channel from coil to MR system needed to send clock information alongside with MR data	62-65
on-coil B_0 shimming	RF coil as B_0 shim element & Wi-Fi transponder	• independency of RF/DC/Wi-Fi operating modes	+ no separate Wi-Fi transponder required – shim current must be provided via battery pack/external power supply	67
wireless power supply				
on-coil power supply	non-magnetic batteries	• available power ≥ power required by on-coil components • low-power on-coil components • wireless power supply integration on-coil/in-bore	+ simple implementation – restricted availability of fully MR compatible batteries – possible image artifacts – trade off between available power capacity and battery size and weight – need for recharging limits scan time	21, 68, 69
	energy harvesting		+ uses RF & gradient energy available during MRI scans – low power supply (≈ mW range)	74-77
	RF WPT		+ high power supply (< 13 W) compared to other options + negligible impact on imaging performance demonstrated – low WPT distance (few cm) – dedicated WPT system needed in-bore	78, 79
	optical WPT		+ optical signal immune to RF – MR compatibility of components to be tested – low demonstrated power supply capability (≈ mW range) – direct line-of-sight required – eye-safety depending on optical power and wavelength	70, 71

FIGURE 2 | Summary of the state of the art in wireless radio frequency (RF) coil development. Existing technologies/strategies for each subsystem (i.e., MR receive signal chain, control signaling, and on-coil power supply) are analyzed listing specific requirements, benefits, as well as limitations or challenges encountered in their current development.

6.02 [28]) to correctly quantize analog MR signal amplitudes. To date, commercially available high-speed ADCs dedicated to MRI are limited to 16-bit [32, 33], insufficient for some imaging scenarios with very high DR. Concerning the sampling rate, one possibility is direct sampling at the Nyquist rate, employing ADCs capable of sampling at high rates greater than twice the Larmor frequency [34]. However, the essential imaging information of the MR signal lies only within a small signal bandwidth (maximum 1–2 MHz), determined by the maximum gradient strength and the field of view (FOV), modulated onto the carrier wave at the Larmor frequency. Therefore, demodulation of the amplified analog RF signal to baseband (around zero frequency) or to an intermediate frequency (IF) by mixing with a local oscillator (LO) signal on-coil before conversion to digital data is possible. This significantly lowers the ADC sampling rate requirement. Analog down-conversion is often used in traditional systems [31, 35] but can also be advantageous for easier system reconfiguration to other B_0 fields and higher power efficiency (<240 mW/channel [21]). This was shown with broadband on-coil receivers for optical fiber transmission of digital signals from two [21] or four [24] wrist coil channels at 1.5–10.5 T. Direct undersampling corresponds to sampling at lower than twice the maximum frequency and digital demodulation at the same time. This technique was applied for single receive elements at 0.18 and 4.7 T [36, 37]. Multi-channel scalable solutions in combination with optical fibers were proposed for in-field receivers with one ADC per coil element at 1.5 and 3 T [18, 19], and four-channel ADCs for MRI up to 2.4 T with an eight-channel coil [38, 39]. Recent research also demonstrated a digital RF front end adaptable for 16 channels and useable from 1.5 to 11.7 T [20, 22]. Direct (under)sampling approaches are useful, as no analog conversion step is needed prior to digitization, and the amount of on-coil components is usually low. However, this technique can be demanding in terms of power consumption (>1 W/channel [20, 22]). Care has to be taken to remove signal ambiguities, e.g., by quadrature (I/Q) demodulation and digitization method-dependent signal filtering. Using I/Q demodulation, the number of components (e.g., amplifiers, ADCs, filters) after the quadrature mixer will be doubled, as there are two separate (I/Q) signal paths. Therefore, especially with discrete components, the form factor and power consumption of the receiver increase. Nevertheless, it can be advantageous to use baseband (I/Q) demodulation, e.g., in an integrated-circuit (IC) design [21, 23], to keep the resulting data rate at a minimum, which can be lower than with IF conversion or direct (under)sampling approaches.

Data Rate

Taken together, the required data rate for wireless transmission depends on the digitization approach and ADC bit resolution for any MR receive system with a specific B_0 field strength, imaging bandwidth, and number of coil elements. Sequence parameters, such as the receive duty cycle (the ratio between acquisition and repetition time), also influence the effective data rate. Estimations of up to 2.6 Gbps, assuming two coil elements at 1.5 T with direct sampling (130 Msps, 20-bit, 50% receive duty cycle) or 64 coil elements at 7 T with baseband sampling of a high bandwidth signal (2 Msps, 20-bit I/Q, 50% receive duty cycle), reveal that these high resulting data rates are difficult to handle with current wireless technologies, as will be detailed in section "Wireless Transmission Technologies and Protocols".

An evident remedy against high data rate and storage requirements is data compression, which can be realized in the analog domain by means of down-conversion before digitization as described above and/or in the digital domain, which requires dedicated signal processing units on-coil (e.g., a field-programmable gate array (FPGA) and digital frequency synthesizer [40]). Digital strategies for DR compression and coil-wise demodulation can be combined to efficiently reduce the data size to one-third of the original amount [41]. Nonetheless, with digital compression directly after digitization, the number of components and, therefore, also the power needed on-coil will increase [22, 42].

To give an estimate for the minimum data rate requirement, we take a modern clinical MRI setup at 3 T with 64 RF receive elements as a reference. In this case, a data rate of at least 512 Mbps would be desirable, assuming moderate signal bandwidth (500 ksps minimum sampling rate), average DR of around 90 dB (covered by 16-bit I/Q ADCs), 50% receive duty cycle, and baseband demodulation to keep the resulting data rate and component power consumption low. Our estimation is in line with other published values [43, 44], only differing in terms of assumed ADC bit resolution, receive duty cycle, or number of receive elements.

Wireless Transmission Technologies and Protocols

Wireless transmission setups have been investigated for their usability in MRI, testing only the wireless link with "synthetic" MR image data without RF coil or signal conditioning components. Except early work on analog wireless MR signal transmission with carriers in the low gigahertz range (<3 GHz [45–47]), research was mostly oriented toward digital wireless MR signal transceivers following IEEE Wi-Fi standards. For digital wireless communication in MRI, apart from achievable data rate and power consumption, lossless spatial transmission is an important criterion. First MR data transfer tests based on the 802.11b [48] or 802.11n [49] standards revealed that long range (>10 m) comes at the cost of low achievable data rates as well as large and power-consuming antennas. These approaches are clearly impractical for wireless MRI. More recent attempts were conducted with higher carriers in the 5 GHz band (802.11ac Wi-Fi protocol), showing reliable in-bore operation of client and router antennae during an MRI scan at data rates around 90 Mbps [44]. This Wi-Fi approach is interesting as small client routers, used in most portable devices nowadays, are available, providing sufficient spatial range for MRI. Efficient data throughput could be improved up to 350 Mbps, suitable for low-channel and low-bandwidth MRI. However, power consumption for only one transmitter antenna can exceed 1 W [50], which can be problematic with limited wireless on-coil power supply, as explained in section "On-Coil Power Supply". Aiming for enhanced data rate capability and reduced power consumption, subsequent work focused on even higher carriers—60 GHz "WiGig" links—included in the

802.11ad Wi-Fi protocol. At 1.5 T, without the presence of RF pulses or gradients, data rates up to 500 Mbps over 10–65 cm were achieved using a miniature transceiver that can achieve up to 2.5 Gbps with only 14 mW DC power per wireless transmitter [43]. Recently, out-of-bore experiments with shielded WiGig dongles [51] have shown transmission rates of 187–665 Mbps over 3–5.5 m distance. This Wi-Fi standard meets our estimated minimum data rate requirement for a modern clinical MRI setup and is therefore viable for wireless coil arrays. Also, the shorter spatial transmission range of one of the presented 60 GHz links [43] is sufficient for some transceiver positioning variants (see **Figure 1B**).

Optical wireless communication (OWC) [52, 53] with visible, infrared, or ultraviolet light carriers (i.e., several 100 THz) could be an attractive alternative to Wi-Fi with distinct benefits [54]: large license-free bandwidth, small and low-power components, immunity to electromagnetic interference, and the possibility for integration into available illumination infrastructure; moreover, OWC can operate well below light intensities considered dangerous for the human eye. Data rates over 3 Gb/s in visible light communication have been shown using a single LED [55]. An MR-compatible OWC front end has been tested for 2 m analog positron emission tomography detector signal transmission [56], but the technology has not yet been exploited for MR signals. Unlike Wi-Fi, high-speed OWC mostly requires a direct line of sight between transceivers, although some systems can even communicate via diffuse light reflections [57]. Suitable components for Li-Fi (Light-Fidelity, i.e., high-speed optical wireless networking [58]) in MRI still remain to be identified and tested on-coil in future studies.

Authors' Opinion on a Wireless MR Receive Signal Chain

Wireless digital MR signal transmission appears feasible with current Wi-Fi strategies under the condition that appropriate measures for data rate reduction prior to wireless transmission are implemented on-coil, e.g., analog baseband demodulation, if possible even combined with further digital data compression methods. Wi-Fi protocol-dependent or component-related drawbacks, e.g., the trade-off between achievable data rate, spatial transmission range, and required power as well as questionable full MR compatibility, restrict the usability of today's Wi-Fi technologies. WiGig (60 GHz) seems to be a promising strategy because of high data rate capability, sufficient transmission range, and low power consumption, although full functioning of WiGig hardware on-coil during an MR scan and the effect on image quality still have to be examined. Also, the final interfacing of the chosen wireless (WiGig) transceiver to a digital RF coil still has to be demonstrated and can be challenging, as it requires the smooth interaction of various on-coil components. So far, Wi-Fi technology benefited from rapid development pushed by the portable device industry; therefore, we think that the implementation of future high-performance Wi-Fi transceivers in RF coils is an aspect to be followed up by the research community. Alternatively, OWC strategies could be investigated for wireless MR signal transmission. With OWC, the wireless transmission of uncompressed, directly digitized MR signals could be envisioned, which is advantageous with respect to miniaturized device size and low system complexity but is questionable concerning a limited on-coil power budget.

Control Signaling

Striving for full removal of coil cabling, a bidirectional wireless link is indispensable as signals must be sent not only from the coil to the MR scanner but also from the scanner control unit to the coil, mainly for triggering, synchronization, and in some cases, control of B_0 shimming.

Active Detuning

Trigger signals need to be distributed to the coil electronics, e.g., to bias PIN diodes for detuning receive coils during RF transmission. Wireless detuning triggers transmitted via a 418 MHz antenna during an MRI scan at 1.5 T have been investigated [59], involving power-efficient replacement of PIN diodes by field-effect transistors (FETs) [60]. Presumably, these trigger signals could also be applied to activate power-consuming components (preamplifiers, ADCs) only during signal reception.

Synchronization

A stable clock, phase-synchronous with the MRI, controlling on-coil electronics (such as ADC or down-conversion), is critical. Clock jitter, which decreases the effective number of ADC bits and creates image artifacts, must be limited. For synchronization of MR unit and in-bore receivers, one method is to physically transmit the MRI master clock to the receiver, which has been demonstrated with 1.6, 2.4, and 3.5 GHz carriers [34, 61]. This requires additional on-coil clocking electronics (e.g., a phase-locked loop, PLL) and a wireless back channel from the MR unit to the coil. In contrast, on-coil clock generators can be used but are particularly impaired by gradient induction; therefore, free-running oscillator information has to be sent to the MR system alongside sampled data to detect and correct for frequency and phase errors as well as time offsets, i.e., to synchronize the two clocks by software. This often requires both additional hardware and software in the wireless receive system [62–65].

On-Coil B_0 Shimming

Several MRI applications benefit from localized on-coil B_0 shimming with DC currents on the RF coil elements compensating for B_0 inhomogeneities [66]. High shim currents themselves cannot be wirelessly transmitted but can be wirelessly controlled, which has been successfully demonstrated by 2.4 GHz Wi-Fi communication [67], using the RF coil itself as a wireless transponder.

Authors' Opinion on Wireless Control Signaling

Overall, less stringent requirements concerning data rate and DR apply to wireless control signals, but correct timing and reliable, simultaneous operation to other wireless paths, especially the MR signal transmission, play a crucial role. Wireless control of active detuning and on-coil B_0 shimming circuits is feasible with existing technologies and has been implemented during an MRI scan in combination with a wired or battery power supply and MR signal transmission via coaxial cables. Solutions

for the synchronization of LO signals or ADC sampling clocks to the MR system clock, crucial to avoid image artifacts and signal degradation, were presented but not demonstrated with a realistic wireless MR receive chain yet because implementation in practice seems challenging. Patient movement and coil vibrations can become an issue for synchronization, but to date, physical system clock transmission via a wireless back-channel appears to be a quite robust solution for wireless MRI. The long-term stability of the external reference clock might be combined with further clock correction in post-processing. Also, a possibility for software synchronization with a free-running oscillator might be included in any case as a fallback strategy if the physical clock transmission fails.

On-Coil Power Supply

The electric power required on-coil is of major concern for wireless RF coil development. In wired coils, generally only components for preamplification and detuning (plus B_0 shimming in some applications) have to be supplied with DC power. In contrast, wireless digital MR signal transmission will add on-coil power requirements for ADCs, potential down-conversion, and wireless transceivers. In this case, the power budget can easily exceed 1–2 W per channel, especially with high-speed ADCs. Power requirements scale with the number of receive channels and depend on multiplexing strategies, i.e., if one ADC and/or wireless transceiver is used for one or multiple coil element(s). For a 64-channel coil and one direct sampling ADC per channel, the power requirement could thus exceed 100 W, which is not feasible with current wireless power supply strategies in MRI as detailed below. Therefore, the first step to implement power supply for wireless coils is to reduce power consumption. Realizable low-power solutions for digitization, detuning, and wireless transceivers have been investigated in studies cited above [21, 43, 60] and could be further improved employing passive components whenever possible, e.g., passive mixers for down-conversion. Assuming a low power consumption in the range of hundreds of milliwatts per receive channel, for arrays up to 64 channels, this still results in on-coil power requirements of tens of watts.

Batteries

The use of non-magnetic rechargeable batteries could be envisioned, although available battery power capacities are limited, and as a consequence, the need for recharging limits scan time. Li-ion batteries (e.g., 5,000 mAh, 7.2 V [21]) or, more specifically, Lithium-ion polymer batteries, e.g., used for motion sensors (250 mAh, 3.7 V, $6.5 \times 18 \times 25$ mm^3 [68, 69]), themselves are generally non-magnetic. However, care must be taken because voltage conversion circuits often include ferrite core transformers not suitable for use in MRI. Typically, an increase in power capacity means bigger battery pack size (e.g., 6,000 mAh, 3.7 V, $5.8 \times 58 \times 138$ mm^3 [69]), and it is therefore obvious that with higher channel count, battery power supply becomes cumbersome and suboptimal for use in-bore or on-coil with limited space.

Wireless Power Transfer

Optical wireless power transfer (WPT) has been suggested for recharging medical implants (<10 mW [70]) or portable devices [71] and could be used in analogy to power-over-fiber approaches previously employed in MRI [12, 72]. To satisfy the power budget for an MR receiver array, it is likely that multiple free-space lasers with high optical powers in combination with efficient photodetectors would be required, possibly resulting in solutions that—depending on optical powers and wavelengths—are not eye-safe [73] and would require sophisticated alignment mechanisms.

For MRI, WPT in the RF range and energy harvesting have been investigated as attractive alternatives. The latter converts energy from electromagnetic fields present during an MR examination, namely the transmit RF field (tens of kilowatts) and gradient fields, into DC power, using inductive coupling in resonant "harvesting" loops [74–77]. Harvesting loops rely on induction at the Larmor frequency, and thus, to avoid system interferences, the size and placement of the loops cannot be chosen freely; further, variations in harvested power depending on the imaging sequence have to be taken into account, which limits the achievable power supply (tens of milliwatts). RF WPT implies the construction of a dedicated system consisting of primary (e.g., in the patient table) and secondary (close to the receive coil) loops for the sole purpose of power delivery by inductive coupling. Byron et al. [78, 79] propose an MR-compatible WPT system operating at 10 MHz transferring up to 13 W over a few centimeters' distance in a 1.5 T system.

Authors' Opinion on Wireless On-Coil Power Supply

The analysis of existing approaches for wireless on-coil power supply leads us to the conclusion that this aspect is still a bottleneck, currently preventing completely wireless MRI. Limitations due to available on-coil power reappear in every subsystem, e.g., concerning the choice of digitization components, analog/digital compression steps, and wireless transceivers. To overcome this bottleneck, ideally, solutions should be found to reduce the total power consumption per wireless MR channel to around 200 mW, so that a 13 W RF WPT system would be sufficient to supply DC power for a 64-channel coil array. Further advances in wireless power development are also desirable to increase available on-coil power budget and therefore alleviate related restrictions. Batteries are currently the only solution for a simple implementation of on-coil power supply, but considering weight, size, and uncertain MR compatibility in some cases, this approach should not remain the only accessible strategy in the future. Out of the other existing strategies, we believe that RF WPT is currently the most sophisticated and promising wireless power supply solution for receive arrays including electronics, as it is capable of supplying a high amount of DC power with negligible impact on MRI performance. A drawback of RF WPT is that the developed system is not yet optimized for on-coil (secondary loop) or in-bore (primary loop) integration. Power transfer distance should ideally be increased and system size and complexity reduced to yield an easily reproducible and efficient WPT solution. Perhaps, another alternative DC power source in MRI

might be a technology based on the magnetoelectric effect, using a piezoelectric material between magnetostrictive layers [80]. However, this technology has not been adapted for MRI conditions yet and will, as we believe, rather be suitable for power delivery in the milliwatt range, similar to existing harvesting techniques, as it is now employed for medical implant charging.

DISCUSSION AND CONCLUSION

In this paper, we summarized the status quo of wireless RF coil development and analyzed existing strategies for the adaptation of the three subsystems of wireless RF coils: the MR receive signal chain, control signaling, and on-coil power supply. We reviewed the benefits of current technology as well as technological challenges or limitations encountered in their development and suggest some future directives.

Over the last years, considerable progress has been made investigating wireless MR and control signal transmission. Feasible strategies exist for on-coil digitization, wireless in-bore signal transmission, cordless active detuning, synchronization to the MR system, and B_0 shim control. However, regarding the numerous requirements for a complete removal of coil cabling in high-density coil arrays, there is still a need for improvement. Solutions described in this work have limitations concerning data rate capabilities and spatial transmission distance as well as power consumption and device size. In addition, full MR compatibility is often questionable. Despite required innovations, we think that future work should focus on the first demonstration of a complete bidirectional wireless MR and control signal chain. This implies the connection of an RF coil with on-coil digitization to a suitable wireless transceiver and the inclusion of wireless active detuning and synchronization circuitry on-coil (leaving out B_0 shimming in a first step, reserved for some specific applications). An important aspect is to thoroughly test this assembly under realistic MRI scan conditions, i.e., with B_0, RF, and gradient fields present and patient movement or coil vibrations possibly impairing component functioning, especially wireless links, and MRI performance. For a proof of concept, only a low number of RF receive elements could be targeted to circumvent high system complexity and high demands in terms of system miniaturization, required data rate, and on-coil power.

Already with low channel counts, wireless on-coil power supply seems to be the main bottleneck, currently preventing fully wireless MRI. Other than bulky rechargeable batteries, no easily accessible WPT technology exists. We believe that reduction of on-coil component power consumption will be achieved and more efficient technologies for WPT will be developed that can be more easily integrated in existing MR systems.

In conclusion, based on our investigations of the state of the art, we predict that completely wireless RF coils will be feasible in the future. Their final implementation will require the combination of already-available technologies and the investigation of alternative promising strategies. Ultimately, with innovations especially required for wireless technologies (e.g., OWC for MRI), MR-compatible components, as well as wireless power supply, efficient solutions for each of the subsystems could be assembled. The realization of wireless RF coils would lead to a significant improvement in coil usability, image quality, patient safety, and comfort.

In the future, wireless RF coils could also follow the trend of additional sensor integration, providing a multitude of complementary information during MRI, e.g., patient motion [81, 82], to further improve image quality and physiological monitoring. While wireless sensor data transmission often relaxes data rate constraints, efficient power supply and reliable data transmission still have to be ensured.

AUTHOR CONTRIBUTIONS

RF-K initiated the work. LN, RF-K, EL, and J-CG contributed to the literature search. LN, RF-K, and EL generated the figures. All authors contributed to writing and proofreading the manuscript.

FUNDING

This work was funded by the Austrian/French FWF (Austrian Science Fund)/ANR grant, No. I-3618 BRACOIL, and the Austrian/French OeAD WTZ grant FR 03/2018.

REFERENCES

1. Moser E. Ultra-high-field magnetic resonance: why and when? *World J Radiol.* (2010) **2**:37–40. doi: 10.4329/wjr.v2.i1.37
2. Moser E, Laistler E, Schmitt F, Kontaxis G. Ultra-high field NMR and MRI—the role of magnet technology to increase sensitivity and specificity. *Front Phys.* (2017) **5**:33. doi: 10.3389/fphy.2017.00033
3. Roemer PB, Edelstein WA, Hayes CE, Souza SP, Mueller OM. The NMR phased array. *Magn Reson Med.* (1990) **16**:192–225. doi: 10.1002/mrm.1910160203
4. Sodickson DK, Manning WJ. Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. *Magn Reson Med.* (1997) **38**:591–603. doi: 10.1002/mrm.1910380414
5. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Magn Reson Med.* (1999) **42**:952–62.
6. Konings MK, Bartels LW, Smits HFM, Bakker CJG. Heating around intravascular guidewires by resonating RF waves. *J Magn Reson Imaging.* (2000) **12**:79–85. doi: 10.1002/1522-2586(200007)12:1<79::aid-jmri9>3.0.co;2-t
7. Armenean C, Perrin E, Armenean M, Beuf O, Pilleul F, Saint-Jalmes H. RF-induced temperature elevation along metallic wires in clinical magnetic resonance imaging: influence of diameter and length. *Magn Reson Med.* (2004) **52**:1200–6. doi: 10.1002/mrm.20246
8. International Electrotechnical Commission (IEC). *International Standards. Medical Electrical Equipment – Part 2-33: Particular Requirements for the Basic Safety and Essential Performance of Magnetic Resonance Equipment for Medical Diagnosis (IEC-60601-2-33). 3.1.* Geneva (2013).
9. Peterson DM, Beck BL, Duensing GR, Fitzsimmons JR. Common mode signal rejection methods for MRI: reduction of cable shield currents for high static magnetic field systems. *Concepts Magn Reson Part B Magn Reson Eng.* (2003) **19**:1–8. doi: 10.1002/cmr.b.10090
10. Seeber DA, Jevtic J, Menon A. Floating shield current suppression trap. *Concepts Magn Reson Part B Magn Reson Eng.* (2004) **21**:26–31. doi: 10.1002/cmr.b.20008

11. Yuan J, Wei J, Shen GX. A 4-channel coil array interconnection by analog direct modulation optical link for 1.5-T MRI. *IEEE Trans Med Imaging.* (2008) **27**:1432–8. doi: 10.1109/TMI.2008.922186
12. Memis OG, Eryaman Y, Aytur O, Atalar E. Miniaturized fiber-optic transmission system for MRI signals. *Magn Reson Med.* (2008) **59**:165–73. doi: 10.1002/mrm.21462
13. Fandrey S, Weiss S, Müller J. A novel active MR probe using a miniaturized optical link for a 1.5-T MRI scanner. *Magn Reson Med.* (2012) **67**:148–55. doi: 10.1002/mrm.23002
14. Koste GP, Nielsen MC, Tolliver TR, Frey RL, Watkins RD. Optical MR receive coil array interconnect. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 13*. Miami Beach (2005). p. 411.
15. Biber S, Baureis P, Bollenbeck J, Höcht P, Fischer H. Analog optical transmission of 4 MRI receive channels with high dynamic range over one single optical fiber. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 16*. Toronto (2008). p. 1120.
16. Demir T, Delabarre L, Akin B, Adriany G, Ugurbil K, Atalar E. Optical transmission system for high field systems. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 19*. Montréal (2011). p. 1865.
17. Du C, Yuan J, Shen GX. Comparison of FP, VCSEL and DFB laser diode in optical transmission for MR RF coil array. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 15*. Berlin (2007). p. 1041.
18. Possanzini C, Van Liere P, Roeven H, Den Boef J, Saylor C, Van Eggermond J, et al. Scalability and channel independency of the digital broadband dStream architecture. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 19*. Montréal (2011). p. 5103.
19. Possanzini C, Harvey PR, Ham K, Hoogeveen R. *The Digital Revolution in MRI With dStream Architecture.* (2016) Available online at: http://clinical.netforum.healthcare.philips.com/global/Explore/White-Papers/MRI/Ingenia-dStream-architecture-the-digital-revolution-in-MRI (accessed September 2, 2019).
20. Reber J, Marjanovic J, Brunner DO, Port A, Pruessmann KP. Scalable, in-bore array receiver platform for MRI. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 24*. Singapore (2016). p. 2170.
21. Sporrer B, Wu L, Bettini L, Vogt C, Reber J, Marjanovic J, et al. A fully integrated dual-channel on-coil CMOS receiver for array coils in 1.5–10.5 T MRI. *IEEE Trans Biomed Circuits Syst.* (2017) **11**:1245–55. doi: 10.1109/TBCAS.2017.2764443
22. Reber J, Marjanovic J, Brunner DO, Port A, Schmid T, Dietrich BE, et al. An in-bore receiver for magnetic resonance imaging. *IEEE Trans Med Imaging.* (2019). doi: 10.1109/TMI.2019.2939090. [Epub ahead of print].
23. Sporrer B, Bettini L, Vogt C, Mehmann A, Reber J, Marjanovic J, et al. Integrated CMOS receiver for wearable coil arrays in MRI applications. In: *Design, Automation & Test in Europe Conference & Exhibition (DATE)*. (2015) p. 1689–94. doi: 10.7873/DATE.2015.1152
24. Port A, Reber J, Vogt C, Marjanovic J, Sporrer B, Wu L, et al. Towards wearable MR detection: a stretchable wrist array with on-body digitization. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 26*. Paris (2018). p. 17.
25. Frass-Kriegl R, Navarro de Lara LI, Pichler M, Sieg J, Moser E, Windischberger C, et al. Flexible 23-channel coil array for high-resolution magnetic resonance imaging at 3 Tesla. *PLoS ONE.* (2018) **13**:e0206963. doi: 10.1371/journal.pone.0206963
26. Mehmann A, Varga M, Vogt C, Port A, Reber J, Marjanovic J, et al. On the bending and stretching of liquid metal receive coils for magnetic resonance imaging. *IEEE Trans Biomed Eng.* (2019) **66**:1542–8. doi: 10.1109/TBME.2018.2875436
27. De Zanche N. MR receive chain. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 27*. Montréal (2019).
28. Gabr RE, Schär M, Edelstein AD, Kraitchman DL, Bottomley PA, Edelstein WA. MRI dynamic range and its compatibility with signal transmission media. *J Magn Reson.* (2009) **2**:137–45. doi: 10.1016/j.jmr.2009.01.037
29. Behin R, Bishop J, Henkelman RM. Dynamic range requirements for MRI. *Concepts Magn Reson Part B Magn Reson Eng.* (2005) **26**:28–35. doi: 10.1002/cmr.b.20042
30. Yuan J, Wei J, Du C, Shen GX. Investigation of dynamic range requirement for MRI signal transmission by optical fiber link. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 15*. Berlin (2007). p. 995.
31. Hashimoto S, Kose K, Haishi T. Comparison of analog and digital transceiver systems for MR imaging. *Magn Reson Med Sci.* (2014) **13**:285–91. doi: 10.2463/mrms.2013-0114
32. Texas Instruments. *Magnetic resonance Imaging (MRI).* Available online at: http://www.ti.com/solution/mri_magnetic_resonance_imaging (accessed September 3, 2019).
33. *ADI's Magnetic Resonance Imaging (MRI) Solutions*. Available online at: https://www.analog.com/media/cn/technical-documentation/apm-pdf/adi-mri-solution_en.pdf (accessed September 3, 2019).
34. Sekiguchi T, Akita K, Nakanishi T, Kato S, Adachi K, Okamoto K. Development of digital wireless transceiver for a MRI coil with clock synchronization. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 17*. Honolulu (2009). p. 3048.
35. Bollenbeck J, Vester M, Oppelt R, Kroeckel H, Schnell W. A high performance multi-channel RF receiver for magnet resonance imaging systems. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 13*. Miami Beach (2005). p. 860.
36. Giovannetti G, Hartwig V, Viti V, Gaeta G, Francesconi R, Landini L, et al. Application of undersampling technique for the design of an NMR signals digital receiver. *Concepts Magn Reson Part B Magn Reson Eng.* (2006) **29**:107–14. doi: 10.1002/cmr.b.20065
37. Pérez P, Santos A, Vaquero JJ. Potential use of the undersampling technique in the acquisition of nuclear magnetic resonance signals. *Magn Reson Mater Phys Biol Med.* (2001) **13**:109–17. doi: 10.1016/S1352-8661(01)00137-5
38. Tang W, Sun H, Wang W. A digital receiver module with direct data acquisition for magnetic resonance imaging systems. *Rev Sci Instrum.* (2012) **83**:104701. doi: 10.1063/1.4755089
39. Tang W, Wang W, Liu W, Ma Y, Tang X, Xiao L, et al. A home-built digital optical MRI console using high-speed serial links. *Magn Reson Med.* (2015) **74**:578–88. doi: 10.1002/mrm.25403
40. De Zanche N. MRI technology: circuits and challenges for receiver coil hardware. In: Iniewski K, editor. *Medical Imaging: Principles, Detectors, and Electronics*. Hoboken, NJ: John Wiley & Sons, Inc. (2009). p. 285–301.
41. Jutras JD, Fallone BG, De Zanche N. Efficient multichannel coil data compression: a prospective study for distributed detection in wireless high-density arrays. *Concepts Magn Reson Part B Magn Reson Eng.* (2011) **39**:64–77. doi: 10.1002/cmr.b.20191
42. Marjanovic J, Reber J, Brunner DO, Engel M, Kasper L, Dietrich BE, et al. A reconfigurable platform for magnetic resonance data acquisition and processing. *IEEE Trans Med Imaging.* (2019). doi: 10.1109/TMI.2019.2944696. [Epub ahead of print].
43. Aggarwal K, Joshi KR, Rajavi Y, Taghivand M, Pauly JM, Poon ASY, et al. A millimeter-wave digital link for wireless MRI. *IEEE Trans Med Imaging.* (2017) **36**:574–83. doi: 10.1109/TMI.2016.2622251
44. Vassos C, Robb F, Vasanawala S, Pauly J, Scott G. Characterization of In-Bore 802.11ac Wi-Fi performance. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 27*. Montréal (2019). p. 1543.
45. Scott G, Yu K. Wireless transponders for RF coils: systems issues. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 13*. Miami Beach (2005). p. 330.
46. Heid O, Vester M, Cork P, Hulbert P, Huish DW. Cutting the cord - wireless coils for MRI. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 17*. Honolulu (2009). p. 100.
47. Riffe MJ, Heilman JA, Gudino N, Griswold MA. Using on-board microprocessors to control a wireless MR receiver array. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 17*. Honolulu (2009). p. 2936.
48. Wei J, Liu Z, Chai Z, Yuan J, Lian J, Shen GX. A realization of digital wireless transmission for MRI signals based on 802.11b. *J Magn Reson.* (2007) **186**:358–63. doi: 10.1016/j.jmr.2007.03.003
49. Shen GX, Wei J, Pang Y. Design of digital wireless transmission for 64 channel array using IEEE 802.11n. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 16*. Toronto (2008). p. 1121.
50. Saha SK, Deshpande P, Inamdar PP, Sheshadri RK, Koutsonikolas D. Power-throughput tradeoffs of 802.11n/ac in smartphones. In: *2015 IEEE Conference*

on Computer Communications (INFOCOM). Kowloon (2015). p. 100–8. doi: 10.1109/INFOCOM.2015.7218372
51. Ko Y, Bi W, Felder J, Shah NJ. Wireless digital data transfer based on WiGig/IEEE 802.11ad with self-shielded antenna gain enhancement for MRI. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 27*. Montréal (2019). p. 1537.
52. Barry JR. *Wireless Infrared Communications*. Norwell, MA: Kluwer Academic Publishers (1994).
53. Chu T-S, Gans M. High speed infrared local wireless communication. *IEEE Commun Mag.* (1987) **25**:4–10. doi: 10.1109/MCOM.1987.1093675
54. Hou R, Chen Y, Wu J, Zhang H. A brief survey of optical wireless communication. In: *13th Australasian Symposium on Parallel and Distributed Computing (AusPDC 2015)*. Sydney (2015) p. 41–50.
55. Tsonev D, Chun H, Rajbhandari S, McKendry JJD, Videv S, Gu E, et al. A 3-Gb/s single-LED OFDM-based wireless VLC link using a gallium nitride μLED. *IEEE Photonics Technol Lett.* (2014) **26**:637–40. doi: 10.1109/LPT.2013.2297621
56. Konstantinou G, Ali W, Chil R, Cossu G, Ciaramella E, Vaquero J. Experimental demonstration of an optical wireless MRI compatible PET/SPECT insert front-end. In: *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*. Strasbourg (2016). p. 1–4. doi: 10.1109/NSSMIC.2016.8069524
57. Lu Z, Tian P, Fu H, Montes J, Huang X, Chen H, et al. Experimental demonstration of non-line-of-sight visible light communication with different reflecting materials using a GaN-based micro-LED and modified IEEE 802.11ac. *AIP Adv.* (2018) **8**:105017. doi: 10.1063/1.5048942
58. Tsonev D, Videv S, Haas H. Light fidelity (Li-Fi): towards all-optical networking. In: *Proceedings SPIE 9007, Broadband Access Communication Technologies VIII, 900702*. San Francisco, CA (2014). doi: 10.1117/12.2044649
59. Lu JY, Robb F, Pauly J, Scott G. Wireless Q-spoiling of Receive Coils at 1.5T MRI. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 25*. Honolulu (2017). p. 4297.
60. Lu JY, Grafendorfer T, Zhang T, Vasanawala S, Robb F, Pauly JM, et al. Depletion-mode GaN HEMT Q-spoil switches for MRI coils. *IEEE Trans Med Imaging.* (2016) **35**:2558–67. doi: 10.1109/TMI.2016.2586053
61. Lu JY, Grafendorfer T, Robb F, Winkler S, Vasanawala S, Pauly JM, et al. Clock transmission methods for wireless MRI: a study on clock jitter and impact on data sampling. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 27*. Montréal (2019). p. 1542.
62. Scott G, Robb F, Pauly J, Stang P. Software synchronization of independent receivers by transmit phase tracking. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 25*. Honolulu (2017). p. 4311.
63. Reykowski A, Redder P, Calderon Rico R, Wynn T, Ortiz T, Dowling G, et al. High precision wireless clock recovery for on-coil MRI receivers using round-trip carrier phase tracking. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 26*. Paris (2018). p. 27.
64. Scott G, Vasanawala S, Robb F, Stang P, Pauly J. Pilot tone software synchronization for wireless MRI receivers. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 26*. Paris (2018). p. 25.
65. Reber J, Marjanovic J, Schildknecht C, Brunner DO, Pruessmann KP. Correction of gradient induced clock phase modulation for in-bore sampling receivers. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 25*. Honolulu (2017). p. 1056.
66. Truong T-K, Darnell D, Song AW. Integrated RF/shim coil array for parallel reception and localized B0 shimming in the human brain. *Neuroimage.* (2014) **103**:235–40. doi: 10.1016/j.neuroimage.2014.09.052
67. Darnell D, Cuthbertson J, Robb F, Song AW, Truong TK. Integrated radio-frequency/wireless coil design for simultaneous MR image acquisition and wireless communication. *Magn Reson Med.* (2019) **81**:2176–83. doi: 10.1002/mrm.27513
68. Chen B, Weber N, Odille F, Large-Dessale C, Delmas A, Bonnemains L, et al. Design and validation of a novel MR-compatible sensor for respiratory motion modeling and correction. *IEEE Trans Biomed Eng.* (2017) **64**:123–33. doi: 10.1109/TBME.2016.2549272
69. PowerStream. Available online at: https://www.powerstream.com/non-magnetic-lipo.htm (accessed November 30, 2019).
70. Saha A, Iqbal S, Karmaker M, Fairose Zinnat S, Tanseer Ali M. A wireless optical power system for medical implants using low power near-IR laser. In: *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Seogwipo (2017). doi: 10.1109/EMBC.2017.8037238
71. Wi-Charge. *The Future of Power*. Available online at: https://wi-charge.com/ (accessed September 26, 2019).
72. Werthen JG, Cohen MJ, Wu T-C, Widjaja S. Electrically isolated power delivery for MRI applications. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 14*. Seattle (2006). p. 1353.
73. International Electrotechnical Commission (IEC). *International Standard IEC 60825-1:2014. Safety of Laser Products - Part 1: Equipment Classification and Requirements 3*. Geneva (2014).
74. Riffe MJ, Heilman JA, Griswold MA. Power scavenging circuit for wireless DC power. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 15*. Berlin (2007). p. 3278.
75. Höfflin J, Fischer E, Hennig J, Korvink JG. Energy Harvesting towards autonomous MRI detection. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 21*. Salt Lake City (2013). p. 728.
76. Middelstaedt L, Foerster S, Doebbelin R, Lindemann A. Power electronics for an energy harvesting concept applied to magnetic resonance tomography. In: *Progress in Electromagnetics Research Symposium*. Prague (2015). p. 1419–23.
77. Byron K, Robb F, Vasanawala S, Pauly J, Scott G. Harvesting power wirelessly from MRI scanners. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 27*. Montréal (2019). p. 1535.
78. Byron K, Robb F, Stang P, Vasanawala S, Pauly J, Scott G. An RF-gated wireless power transfer system for wireless MRI receive arrays. *Concepts Magn Reson Part B Magn Reson Eng.* (2017) **47B**:e21360. doi: 10.1002/cmr.b.21360
79. Byron K, Winkler SA, Robb F, Vasanawala S, Pauly J, Scott G. An MRI compatible RF MEMs controlled wireless power transfer system. *IEEE Trans Microw Theory Tech.* (2019) **67**:1717–26. doi: 10.1109/TMTT.2019.2902554
80. Rizzo G, Loyau V, Nocua R, Lourme JC, Lefeuvre E. Potentiality of magnetoelectric composites for wireless power transmission in medical implants. In: *13th International Symposium on Medical Information and Communication Technology (ISMICT)*. Oslo (2019). p. 1–4. doi: 10.1109/ISMICT.2019.8743873
81. Schildknecht CM, Brunner DO, Schmid T, Reber J, Marjanovic J, Pruessmann KP. Wireless motion tracking with short-wave radiofrequency. In: *Proceedings of the International Society for Magnetic Resonance in Medicine 27*. Montréal (2019). p. 66.
82. van Niekerk A, van der Kouwe A, Meintjes E. Toward "plug and play" prospective motion correction for MRI by combining observations of the time varying gradient and static vector fields. *Magn Reson Med.* (2019) **82**:1214–28. doi: 10.1002/mrm.27790

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 15 April 2020
doi: 10.3389/fphy.2020.00092

A Flexible Array for Cardiac ^{31}P MR Spectroscopy at 7 T

Sigrun Roat[1], *Martin Vit*[2], *Stefan Wampl*[1], *Albrecht Ingo Schmid*[1] *and Elmar Laistler*[1]*

[1] *Division MR Physics, Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Vienna, Austria,*
[2] *Institute of Mechatronics and Computer Engineering, Technical University of Liberec, Liberec, Czechia*

Purpose: The simulation optimization and implementation of a flexible ^{31}P transmit/receive coil array, under the geometrical constraint of fitting into the housing of an already existing 12-channel proton array, to enable localized cardiac ^{31}P MRS at 7 T is presented.

Methods: The performance in terms of homogeneity, power and SAR efficiency, and receive benchmark of 32 potential array designs was compared by full wave 3D electromagnetic simulation considering the respective optimal static B_1^+ shims. The design with the best performance was built and compared to a commercially available single loop in simulation and measurement.

Results: Simulation revealed an optimal array design comprising three overlapping elements, each sized 94 × 141 mm^2. Simulation comparison with a single loop coil predicted a performance increase due to increased power efficiency and lower SAR values. This was verified by phantom measurements, where an SNR increase of 46% could be observed for localized ^{31}P spectroscopy in a voxel positioned comparable to an *in vivo* cardiac measurement scenario.

Conclusion: A flexible ^{31}P/^{1}H RF coil array with improved SNR is presented, enabling localized *in vivo* cardiac ^{31}P spectroscopy at 7 T.

Keywords: ^{31}P cardiac MRS, 3D EM simulation, RF coil design, X-nucleus, ultra-high field

OPEN ACCESS

Edited by:
Zhen Cheng,
Stanford University, United States

Reviewed by:
Angelo Galante,
University of L'Aquila, Italy
Michael D. Noseworthy,
McMaster University, Canada

***Correspondence:**
Elmar Laistler
elmar.laistler@meduniwien.ac.at

Specialty section:
This article was submitted to Medical Physics and Imaging, a section of the journal Frontiers in Physics

Received: *29 November 2019*
Accepted: *12 March 2020*
Published: *15 April 2020*

Citation:
Roat S, Vit M, Wampl S, Schmid AI and Laistler E (2020) A Flexible Array for Cardiac ^{31}P MR Spectroscopy at 7 T. Front. Phys. 8:92. doi: 10.3389/fphy.2020.00092

INTRODUCTION

Phosphorus (^{31}P) magnetic resonance spectroscopy (MRS) is known to be a powerful tool in the assessment of cell energy metabolism [1–3]. Coronary heart disease is one of the most common causes of death in the western hemisphere. A common reason for cardiac dysfunction is a deficit of the myocardial metabolism [4–6] for which cardiac ^{31}P MRS is a direct and non-invasive assessment method [7–9]. The relative and absolute concentrations of ATP and PCr, and especially their ratio are strong indicators of cardiac dysfunction [10, 11]. The technique is very specific but suffers from inherently low sensitivity.

The low gyromagnetic ratio and *in vivo* concentration of ^{31}P results in a low intrinsic signal to noise ratio (SNR) (^{31}P-MRS has 100,000 × lower SNR than ^{1}H MRI [12]), which leads to low spatial and temporal resolution. The nuclear magnetization increases proportionally with the main magnetic field strength (B_0) results in significant SNR increase for all nuclei detectable by MR. At ultra-high field (≥ 7 T) the proton (^{1}H) B_1^+ homogeneity becomes more challenging in

larger anatomic regions due to the short wavelength, however, this applies less strongly to ^{31}P MRS because of the lower Larmor frequency. ^{31}P cardiac MRS additionally benefits from increased B_0 since the chemical-shift anisotropy helps to shorten the cardiac T_1 relaxation times for at least PCr and ATP, therefore, increasing SNR [12]. Comparison of spectral quality between 3 T and 7 T cardiac ^{31}P MRS, showed the significant benefit that comes with increasing B_0, i.e., a 2.8 fold increase in PCr SNR [13].

To acquire as much of the theoretically available signal available, optimized RF coils need to be employed. The use of multiple receiving and/or transmitting elements in coil arrays offers SNR [1, 14] and/or acquisition speed advantages on the receive side [15, 16] and—if combined with adequate simulation and pulse design—greatly improved data quality and lower specific absorption rate on the transmit side [17, 18]. Electromagnetic coupling of the RF coil to the tissue and, therefore, SNR, increases when the coil is conformed to the anatomy [19]. Hence, most coils are assembled on anatomically form-fitted rigid housings. However, for applications where large anatomical inter-subject variability is expected, flexible RF coils are favorable to account for this heterogeneity in anatomy. Due to the longer wavelength for ^{31}P it is still possible to use single loop coils for 7 T ^{31}P MRS on the torso, which is the RF coil of choice in most published studies [13, 20, 21]. More elaborate designs include a 2-element Tx/Rx overlap array [9] and combinations of a Tx volume coil with Rx arrays [22–24].

The goal of this study was to design, build and evaluate a dedicated flexible ^{31}P RF coil for phosphorus cardiac MR spectroscopy at 7 T to be integrated into a 12-channel transmission line resonator (TLR) array for torso MRI [25]. Due to the already existing coil housing, possible design dimensions for the ^{31}P array were limited. Suitable designs with various element sizes and arrangements were investigated via 3D electromagnetic simulation in a comprehensive study. By definition of a performance measure that takes into account power efficiency, SAR efficiency, and homogeneity of the resulting transmit field B_1^+, and the receive performance based on the resulting B_1^- field, the best performing design was identified and eventually realized. A novel concept for floating dual-tuned cable traps working at both frequencies of operation (297.2 MHz for ^{1}H and 120.3 MHz for ^{31}P) was developed and integrated into the coil housings. The performance of the proposed array was compared with a commercially available standard single loop ^{31}P RF coil for cardiac applications in simulation and measurement. Finally, the feasibility of acquiring localized ^{31}P spectra *in vitro* was demonstrated.

MATERIALS AND METHODS

RF Array Design

Potential RF coil designs were intended to cover the average human heart size of $12 \times 8 \times 6$ cm^3 and its location ∼2 cm below the sternum [26]. The developed ^{31}P array acts as an extension to an existing ^{1}H RF coil [25], to enable acquisition of additional metabolic information of the heart muscle. The proton coil array consists of 12 TLR elements that were fabricated on a flexible substrate with a rigid PCB part in the center of each TLR element connected to their tuning and matching components. The PCB is connected to a rigid housing box incorporating each elements interface board, including T/R switches, 1:3 splitters and cable traps. The considered designs are to fit into the rigid housing boxes of the TLR elements, which poses a hard constraint on the maximum number of elements, coil sizes, and shapes. **Figure 1** shows all considered RF coil array configurations, ranging from 1- to 4-channel arrays differing in size, arrangement and position, yielding a total number of 32 simulated array designs. The 12 TLR elements and their respective shields are depicted in gray. To discretely sample the possible configurations, the element size was varied in multiples of the ^{1}H TLR element dimensions of 94 $\times$ 94 mm^2. Regarding the position, the respective array center matches either the ^{1}H array's center (corresponding to the center of the body) or is shifted by one half ^{1}H element width to the patient's left. The flexibility of the ^{1}H coil leads to a bending of the leftmost and rightmost elements. By shifting the center of the ^{31}P coil the array experiences a different degree of bending which has an influence on the overall produced field. Those positions are denoted body-centered (bc) and heart-centered (hc), respectively.

Electromagnetic Simulation

All coil designs were modeled in XFdtd 7.5 (Remcom, State College, PA, USA) using 1 mm thick wire as perfect conductors. The 3D EM simulations and their post-processing were computed on a workstation equipped with 4 GPUs (Tesla C2070, Nvidia, Santa Clara, CA, USA) enabling GPU acceleration, 12 CPUs (Intel® Xeon® X5690, 12 M Cache, 3.46 GHz, 6.40 GT/s Intel® QPI, Santa Clara, CA, USA), and 190 GB RAM. Each coil element was cut into equally long copper stubs connected by a capacitor to limit the electrical length of the coil. Depending on the configuration it was used in, the number of gaps was 8, 6, and 4 for the 1, 3, and 4 element arrays, respectively. This corresponds to stub-lengths between $\approx \lambda/15$ and $\lambda/42$ for the 2×1.5 elements (4 channel array) and 1.5×1 element (1 element array), respectively, preventing any wavelength effects for all presented designs. All capacitors were eventually replaced by 50 Ω voltage sources to enable a fast RF co-simulation approach [27] in ADS (Keysight Technologies, Santa Rosa, CA, USA). All designs were simulated as overlap-decoupled arrays. An overlap factor of 0.86 was used [28]; as this factor only applies to quadratic elements, for non-quadratic elements decoupling was corrected by additional counter-wound inductances (CWI) [29] during RF co-simulation. Realistic loss incorporation was implemented by assigning capacitors their realistic equivalent series resistances by extrapolating an ESR model for the ATC 100 E capacitor series (http://www.atceramics.com/multilayer_capacitors.html). Solder joint losses were modeled as series resistances extrapolated to 120.3 MHz from literature [30]. Counter-wound inductances were modeled lossless since they were solely used to mimic sufficient overlap decoupling. Losses of the power splitter (−0.36 dB/channel) and the transmit/receive switches (−0.8 dB/channel) were measured on the bench and incorporated into the simulation.

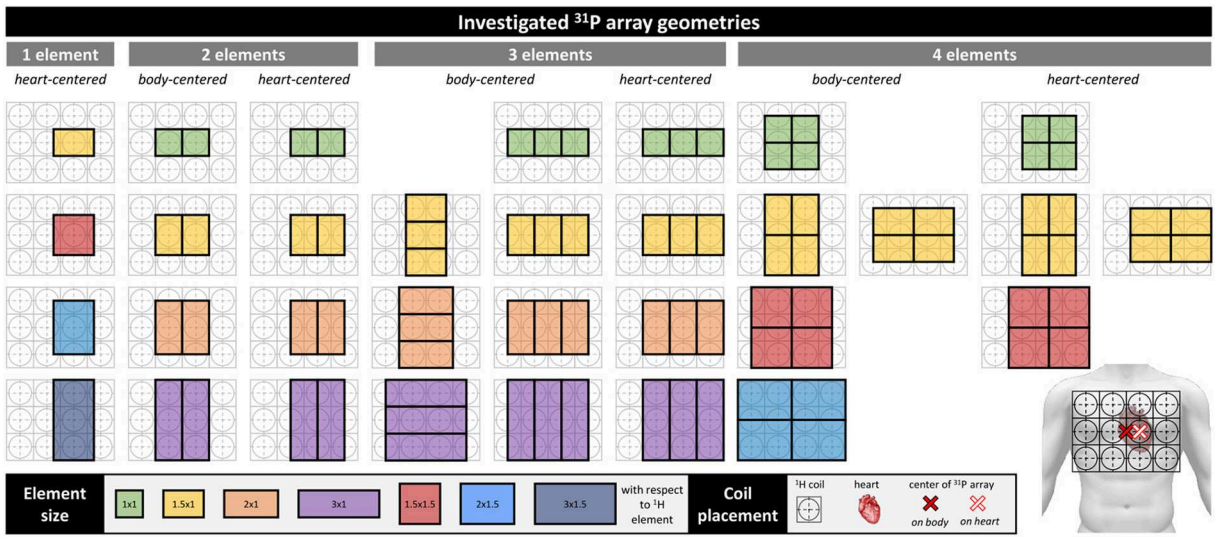


FIGURE 1 | Simulated array designs. The designs are grouped by the number of channels, i.e., 1–4 elements. Each group is divided in body-centered (bc) and heart-centered (hc) arrays. bc-arrays are centered above the sternum, hc-arrays are shifted by half a housing box to the left, to be centered above the heart, as indicated on the torso on the bottom right. Colors represent the size of the individual elements, as stated in the legend on the bottom, the numbers are the element sizes in multiples of the underlying 12-ch ^{1}H array housing boxes (grid on torso, bottom right); one unit cell has a length of 94 mm.

TABLE 1 | The total number of phase sets for each of the four coil groups (1-, 2-, 3-, and 4-element arrays) as well as the total number of simulation setups are stated.

	Total number of phase optimization simulations				
Array design	**1 element**	**2 elements**	**3 elements**	**4 elements**	**Total # phase optimization simulations**
$\Delta\varphi$		5°	5°	10°	
$\lvert\Phi_i\rvert$	1	72	5184	46656	955016
# of distinct designs	4	8	11	9	

Each phase set results in a certain value for PE, RH, SE, and f_φ. The phase set that maximizes f_φ is the designated optimal phase set of the corresponding RF coil. The total number of phase optimization simulations for each voxel model equals 477508.

The proposed array designs were loaded with realistic human body models ("*Duke*" and "*Ella*," Virtual Family, IT'IS Foundation, Zurich, Switzerland), yielding a total of 64 3D simulation setups to be compared. Combination of 3D EM field data and co-simulation results and further post-processing was performed in Matlab 2017b (Mathworks, Natick, MA, USA). In order to compare the performance of the designs, optimal static B_1^+ shimming was obtained by varying the relative phase shift ($\Delta\varphi$) between the elements in 5° steps for the 2- and 3- element arrays and in 10° steps for the 4 element arrays, respectively (see **Table 1** for the total number of phase sets $|\Phi_i|$).

The optimal phases were determined for each design by maximizing a merit function f_φ that is an equally weighted combination of power efficiency (PE), SAR efficiency (SE), and relative homogeneity (RH):

$$\left.\begin{aligned} \mathrm{PE}(\Phi_i) &= \frac{\overline{B_1^+(\Phi_i)}}{\sqrt{P_{in}}} \\ \mathrm{SE}(\Phi_i) &= \frac{B_1^+(\Phi_i)}{\sqrt{\max(SAR_{10g}(\Phi_i))}} \\ \mathrm{RH}(\Phi_i) &= 1 - \frac{\mathrm{std}(B_1^+(\Phi_i))}{B_1^+(\Phi_i)} \end{aligned}\right\}$$

$$f_\varphi(\Phi_i) = \frac{1}{3}\cdot\left(\frac{\mathrm{PE}(\Phi_i)}{\underset{\Phi i}{max}(\mathrm{PE})} + \frac{\mathrm{SE}(\Phi_i)}{\underset{\Phi i}{max}(\mathrm{SE})} + \frac{\mathrm{RH}(\Phi_i)}{\underset{\Phi i}{max}(\mathrm{RH})}\right) \quad (1)$$

In Equation (1) the maximum value is evaluated over all simulated phase combinations for one specific design. The mean values were averaged over an ROI comprising the heart lumen and muscle and normalized with respect to the maximum value for the respective array. To identify the best design (d_i, i = 1,..., 32, i.e., all considered coil designs), an extended merit function f_{tot}, additionally taking into account the receive efficiency in terms of SNR [31] was evaluated for all phase-optimized designs, but now normalized with respect to the maximum values for SE, PE, RH, and SNR over all investigated designs, respectively:

$$\mathrm{SNR} = \frac{|B_1^-|}{\sqrt{P_{abs}}}$$

$$f_{tot}(d_i) = \frac{1}{4}\cdot\left(\frac{\mathrm{PE}(d_i)}{\underset{di}{max}(\mathrm{PE})} + \frac{\mathrm{SE}(d_i)}{\underset{di}{max}(\mathrm{SE})} + \frac{\mathrm{RH}(d_i)}{\underset{di}{max}(\mathrm{RH})} + \frac{\mathrm{SNR}(d_i)}{\underset{di}{max}(\mathrm{SNR})}\right) \quad (2)$$

The values for PE, SE, RH, and SNR were averaged over "*Duke*" and "*Ella*," both equipped with the same design.

To evaluate the influence of the CWI decoupling, another set of simulations with 2 elements of dimensions 1 × 1 ($\hat{=}$ 94 × 94 mm^2) and 1 × 3 ($\hat{=}$ 94 × 282 mm^2) and overlap factors between 0.76 to 0.92 in steps of 0.02 was performed. The arrays were loaded with a rectangular phantom filled with a material mimicking tissue ($\sigma = 0.55$ S/m, $\varepsilon = 51$) and were tuned, matched and decoupled in co-simulation using CWI where necessary. Static B_1^+ shimming for a spherical ROI with a diameter of 125 mm, located 35 mm below the RF coils was derived in the same manner as described in the previous paragraph using Equation (1). The arrays with the overlap factor resulting in best decoupling were compared to the corresponding arrays with overlap factor 0.86 with additional CWI in terms of S_{12}, RH, PE, SE, SNR, and maximum 10 g-SAR. For a theoretical comparison of a commercially available single loop ^{31}P coil (RAPID Biomedical GmbH, Rimpar, Germany) with the best design identified above, the performance of both RF coils was investigated using the aforementioned simulation workflow. The single loop has a diameter of 140 mm and an assumed wire thickness of 1.5 mm. The loss of the transmit/receive switch was set to the same as for the array (−0.8 dB) and incorporated in the evaluation. Both setups were positioned on the voxel models as closely to reality as possible in terms of distance and curvature in order to obtain results comparable to the measured data.

RF Coil Implementation

The design determined by simulation, i.e., the 3 channel 1 × 1.5 heart-centered array, was implemented. For flexibility, the ^{31}P array was constructed out of flexible stranded wire (Ø = 2 mm). Crosstalk between the ^{1}H and ^{31}P arrays was minimized by replacing every second segmenting capacitor of the ^{31}P loops by an LCC trap [32], resulting in three traps per element. To ease the handling of the whole coil and to keep it as flexible as possible, a separate interface box for the coil was avoided by placing transmit-receive switches, preamplifiers and power splitters inside the 12 separated 3D-printed housing boxes of the ^{1}H array. Performance of the ^{31}P array was tested on the bench, measuring S-parameters for five human volunteers (3 male, 2 female) using a vector network analyzer (E5071C, Agilent, Santa Clara, CA, USA).

In order to prevent common mode currents on the cables at both Larmor frequencies, double tuned floating cable traps were implemented [33]. By nesting two floating traps [34] into one another, blocking at two different frequencies can be achieved. Two hollow dielectric cylinders are split in half along their axes and are covered by conductive copper layers on the inside and outside walls. The inner trap shares its outer copper layer with the inner copper layer of the outer trap. At one end, all three concentric copper layers are short-circuited, while tuning capacitors are connecting the outer to the middle and the middle to the inner layer on the other side. The capacitors (CHB series, Exxelia Ceramics, Pessac, France) across the outer (inner) shell coarsely control the first (second) resonance frequency of the

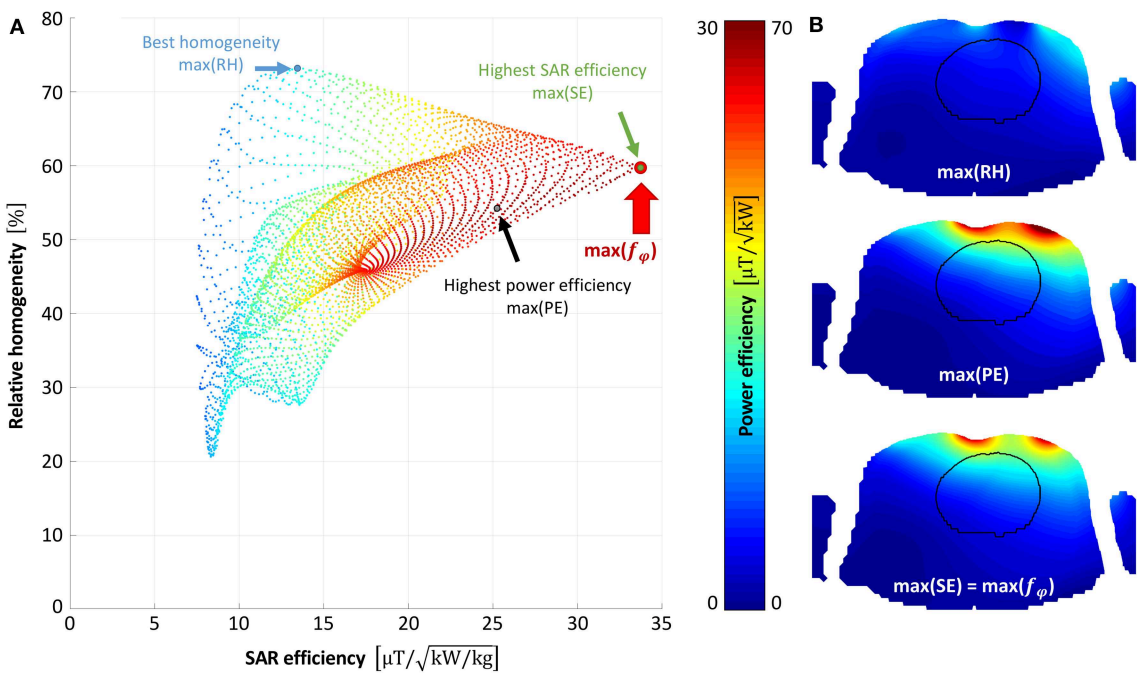


FIGURE 2 | Static B_1^+ shimming result. Static B_1^+ shimming result for the best-performing design (3-element heart centered with 94 × 141 mm² elements) simulated on Duke is shown. **(A)** depicts the resulting point cloud (5,184 points), where each point represents the array's performance in terms of relative homogeneity (RH), power efficiency (PE), and SAR efficiency (SE), for a certain phase set. The optimal phase set is chosen where the merit function f_φ is maximized (depicted as a red circle). In **(B)** the resulting $B_1^+/\sqrt{P_{in}}$ maps are shown for the phase shifts that maximize RH, PE, SE and f, respectively.

trap, respectively. By varying the distance between the two halves of the cylinders, the frequencies can be finely adjusted, however not independently. The trap body was 3D-printed (Rebel 2, Petr Zahradník Computer Laboratory, Ústí nad Labem, Czech Republic) from ABS plastic material ($\varepsilon \approx 2$). The dimensions of the hollow cylinders were Ø = 8 mm, 13 mm, and 20 mm, respectively, all with a length of 55 mm.

MR Measurements

All MR measurements were conducted on a 7 T whole body MR scanner (Siemens Magnetom, Erlangen, Germany), using a tissue mimicking torso phantom with dimensions of 230 × 280 × 380 mm^3 containing saline solution ($\sigma \approx 0.5$ S/m, $\varepsilon \approx 80$) with 1.57 g/L K_2HPO_4 and 0.14 g/L KH_2PO_4 (0.01 M/l PO_4, pH = 8). The coil was positioned centrically with respect to the phantom which was located in the isocenter of the scanner.

^{31}P CSI data was acquired using an ultrashort TE chemical shift imaging (ute-CSI) sequence [35] (TR/TE = 1770/2.3 ms, FOV 400 × 400 × 350 mm^3, matrix size 8 × 16 × 8, vector size 512, scan time 7:46 min). For localized ^{31}P spectroscopy a stimulated echo acquisition mode (STEAM) sequence [36] was applied (TR/TE = 3,000/13.4 ms, TM = 6.9 ms, voxel size 50 × 20 × 50 mm^3, vector size 1024, 32 averages, voxel location 7 cm from phantom wall, scan time 1:36 min). All spectra were post-processed using the MATLAB-based OXford Spectroscopy Analysis (OXSA) toolbox [37] and its implementation of the AMARES fitting method [38]. For combination of the three individual channels of the array a whitened singular value decomposition (WSVD) approach was used [39].

LCC trap performance was validated inside the scanner, by acquiring flip angle maps in an equally sized phantom containing only saline solution ($\sigma \approx 0.5$ S/m, $\varepsilon \approx 80$) with and without the ^{31}P array integrated into the ^{1}H coil housings using a saturated Turbo FLASH (satTFL) sequence [40] (TR/TE = 10,000/2.02 ms, FOV = 450 × 450 mm^2, matrix size 128 × 128, rectangular slice-selective saturation pulse, pulse duration 700 ms, reference voltage 300 V, slice thickness 10 mm).

RESULTS

RF Array Design

The computational cost of each individual full wave simulation depends on the size of the array, the number of gaps, and the voxel model used as load. CPU/GPU RAM requirements and the total computation time for a single EMS were between 0.38/0.25 GB and 1.47 h (design: Ella 1 element 1.5 × 1 hc) and 2.49/1.52 GB and 49.9 h (design: Duke four element 2 × 1.5 bc). Post-processing of the individual designs is highly dependent on the number of channels, ergo the number of different phase sets that need to be calculated in order to evaluate the static B-$_1^+$ shimming, and was between 0.46 and 13.6 min for the Ella 1 element 1.5 × 1 hc and Duke 4 element 2 × 1.5 bc, respectively.

Static B_1^+ shimming was optimized for each of the 64 simulated designs (32 for "*Duke*," and 32 for "*Ella*"). **Figure 2A** shows the resulting point cloud for the B_1^+ shimming procedure for an exemplary dataset. Each point represents the result in terms of RH, PE, and SE for a certain phase set. The phase sets that result in maximum RH, PE, SE, and f_φ, are marked by blue, black, green, and red circles, respectively. **Figure 2B** shows the resulting $B_1^+/\sqrt{P_{\text{in}}}$ maps achieved with the optimal phase set for best RH, best PE, and best SE (which is equal to best f_φ in the shown case). Evaluating the extended merit function f_{tot} over all designs resulted in the final design choice of a 3 element array with element sizes of 94 × 141 mm^2, centered above the heart. **Figure 3** depicts the mean values over the heart ROI for f_{tot}, RH, SE, PE, and SNR for all simulated designs for Ella, Duke, and the average over both (black).

In the 2-channel array simulation for determining the influence of the CWI decoupling, the optimal overlap factor was determined to be 0.88 for the 1 × 1 and 0.78 for the 1 × 3 sized arrays. S_{21} for optimized overlap decoupling (OL) only and fixed overlap plus additional CWI decoupling (OL + CWI) were always below −17.1 db. For both array types, i.e., 1 × 1 and 1 × 3, the highest deviation between OL + CW and OL designs was found in the maximum 10 g SAR value, with an increase of 1.29% ($\hat{=}$ 0.02 1/kg) and 4.13 % ($\hat{=}$ 0.03 1/kg) for the 1 × 1 and 1 × 3 arrays, respectively. All results are summarized in **Table 2**. These findings support the hypothesis that simplifying the simulations of all array designs with a fixed overlap + CWI to mimic optimal overlap decoupling is reasonable.

Bench measurements of the implemented array in loaded condition before and after incorporation into the ^{1}H housing were conducted on 5 human volunteers (3 male, 2 female, 30 ± 3.6 years) and show sufficient matching and isolation between array elements, i.e., the reflection coefficients (S_{11}, S_{22}, S_{33}) were always below −17.5 dB and −17.2 dB, respectively, while transmission coefficients (S_{12}, S_{23}, S_{31}) were below −13.1 and −13.6 dB. The array needed to be slightly retuned and rematched after integration due to slight position changes and distance to the sample. The measured Q ratio (Q_u/Q_l) for all three elements prior and after incorporation was above 5.5 and 5.7, indicating sample loss dominance and negligible additional losses due to the ^{1}H coil and housing. The floating double tuned traps were correctly tuned, with a blocking of −10.5 dB/−34 dB and a bandwidth of 3 MHz/6 MHz at 120 MHz/297.2 MHz respectively. The tuning range for both blocking frequencies by changing the gap size between the half-cylinders was ±10%, which was sufficient to tune the traps to the desired frequencies.

Performance Comparison With Single Loop

In simulation the proposed three element array yields a mean power efficiency in the heart ROI that is 58% higher than the respective values for the single loop reference coil. In terms of SAR efficiency, the array performs 124% better than the loop; the 10 g averaged SAR values decrease by 51%. Relative homogeneity is 30% better. The results are presented in detail in **Table 3** and **Figure 4**.

MR Measurements

A maximum deviation in B_1^+ acquired with and without the ^{31}P array present of <20% was found (see **Figure 5**). Before acquiring CSI data, a series of localized spectra were obtained in order to find the reference voltage for a voxel in a location similar

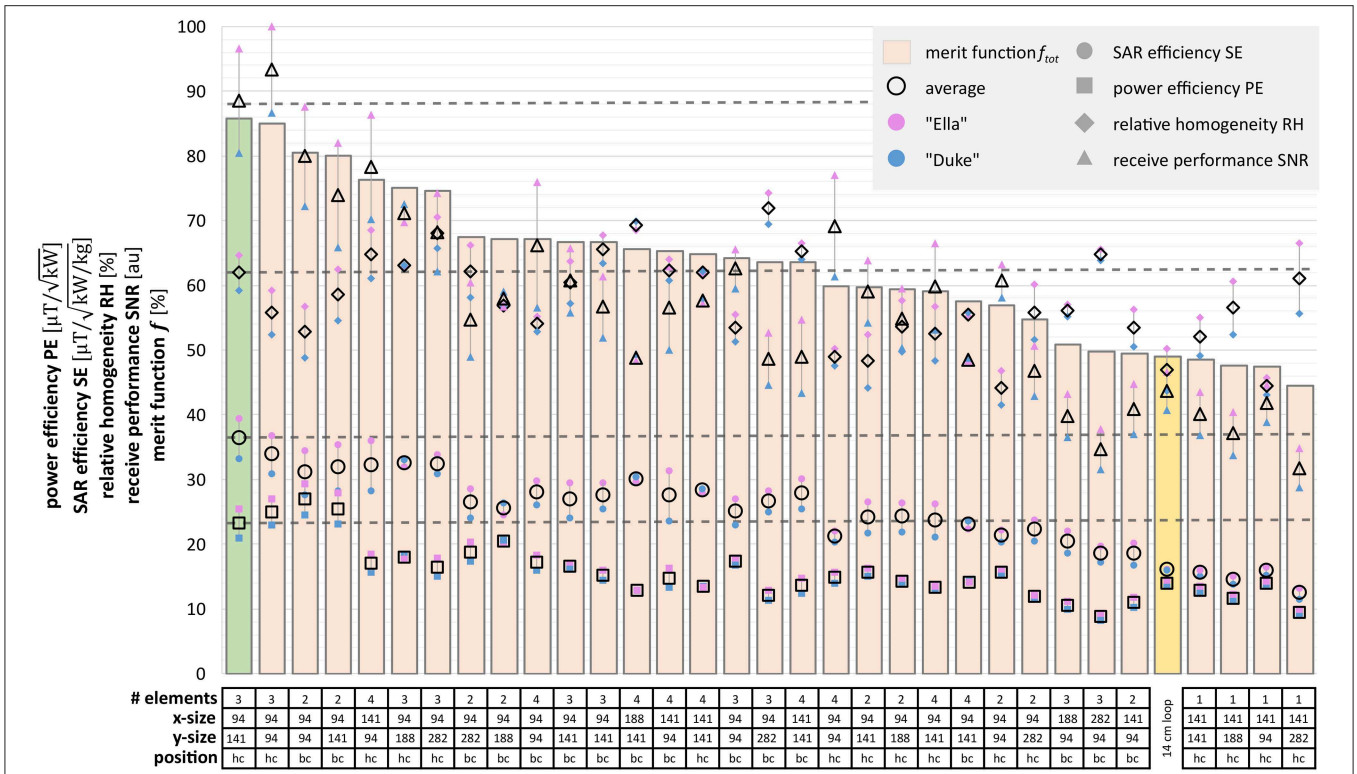


FIGURE 3 | Comparison of the 32 simulated RF coil designs. The performance parameters relative homogeneity (RH, diamonds), power efficiency (PE, squares), SAR efficiency (SE, circles), and receive efficiency (SNR, triangles) are plotted for each design. Results for Ella, Duke, and their average are depicted in purple, blue, and transparent with black rim, respectively. The figure of merit f_{tot} is displayed as bars. The highlighted bar indicates the performance parameters of the reference single loop coil. Below the chart, the table summarizes the geometric properties of each array, x- and y-size are given in multiples of 94 mm, the unit size of the housing boxes of the underlying ^{1}H array. The best design for both Duke and Ella is the 3-element heart-centered array with 94 × 141 mm².

TABLE 2 | Investigation of CWI decoupling elements via simulation of 2 element arrays.

	1 × 1							1 × 3						
	Overlap factor	**S_{12}**	**RH**	**PE**	**SE**	**SNR**	**10g SAR**	**overlap factor**	**S_{12}**	**RH**	**PE**	**SE**	**SNR**	**10 g SAR**
		dB	%	μT/√kW	μT/√(W/kg)	a.u.	1/kg		dB	%	μT/√kW	μT/√(W/kg)	a.u.	1/kg
OL + CWI	0.86	−17.11	27.02	10.24	8.12	3.73	1.59	0.86	−20.06	39.13	7.00	8.17	2.63	0.74
OL	0.88	−17.55	27.28	10.19	8.13	3.73	1.57	0.78	−18.69	37.88	6.88	8.19	2.63	0.71
abs. Difference	0.02	0.44	0.26	0.05	0.01	0.00	0.02	0.08	1.37	1.26	0.12	0.02	0.01	0.03

Arrays with two elements of dimensions 1 × 1 and 1 × 3 were simulated once with an optimized overlap factor (OL) and with a fixed overlap factor of 0.86 and additional counter-wound inductances (OL + CWI). RH, PE, SE, and SNR values are averaged over a spherical ROI volume. Performance difference of the OL + CWI vs. OL arrays can be seen in the bottom row and is negligible for both dimensions, supporting the CWI's use to mimic optimal overlap decoupling.

to the human heart, i.e., 7 cm from phantom surface wall in y-direction. **Figure 6A** shows all spectra plotted as signal amplitude vs. reference voltage. The reference voltage is the voltage that would be required to achieve a 90° flip angle using a 1 ms block pulse. The signal amplitudes were fitted with a $\sin^3$ function, corresponding to the signal equation for STEAM sequences. The reference voltage for the single loop is 400 V, whereas the array needs 880 V in the same voxel. The localized spectra of the acquisitions where 90° were reached are shown in **Figure 6B** for the array and the single loop. From these spectra SNR values of 52 for the array, and 35 for the single loop were calculated. In **Figures 6C,D** transversal and sagittal metabolic maps acquired with the array (top) and the single loop (bottom) are displayed.

DISCUSSION

In this work we show the successful integration of an optimized 3-channel ^{31}P array for cardiac MRS at 7 T into a flexible 12 channel ^{1}H coil.

TABLE 3 | Simulation comparison of 3-element array vs. single loop coil.

		Three-element array					Single loop					% change				
		RH	PE	10g SAR	SE	SNR	RH	PE	10g SAR	SE	SNR	RH	PE	10g SAR	SE	SNR
		%	μT/√kW	1/kg	μT/ √(W/kg)	au	%	μT/√kW	1/kg	μT/ √(W/kg)	au					
Heart	Duke	60.6	20.7	0.37	34.3	10.6	46.4	13.8	0.68	16.8	9.1	30.5	49.7	−46.5	104.8	16.1
	Ella	68.0	24.6	0.38	39.7	12.8	52.7	14.9	0.84	16.3	10.5	29.2	65.2	−54.3	144.2	21.8
	avg.	**64.3**	**22.7**	**0.37**	**37.0**	**11.7**	**49.5**	**14.4**	**0.76**	**16.5**	**9.8**	**29.8**	**57.7**	**−50.8**	**124.2**	**19.2**
VOI	Duke	83.2	13.9	0.37	23.0	8.9	71.9	10.9	0.68	13.2	7.8	15.8	27.1	−46.5	73.9	14.0
	Ella	82.0	16.9	0.38	27.4	9.9	73.9	13.4	0.84	14.7	8.2	10.8	26.4	−54.3	86.8	20.9
	avg.	**82.6**	**15.4**	**0.37**	**25.2**	**9.4**	**72.9**	**12.2**	**0.76**	**13.9**	**8.0**	**13.3**	**26.7**	**−50.8**	**80.7**	**17.5**

The bold values state the averaged values over Duke and Ella for the heart and VOI respectively. RH, PE, SE, and SNR values are averaged over the whole heart volume and over the VOI used in measurement.

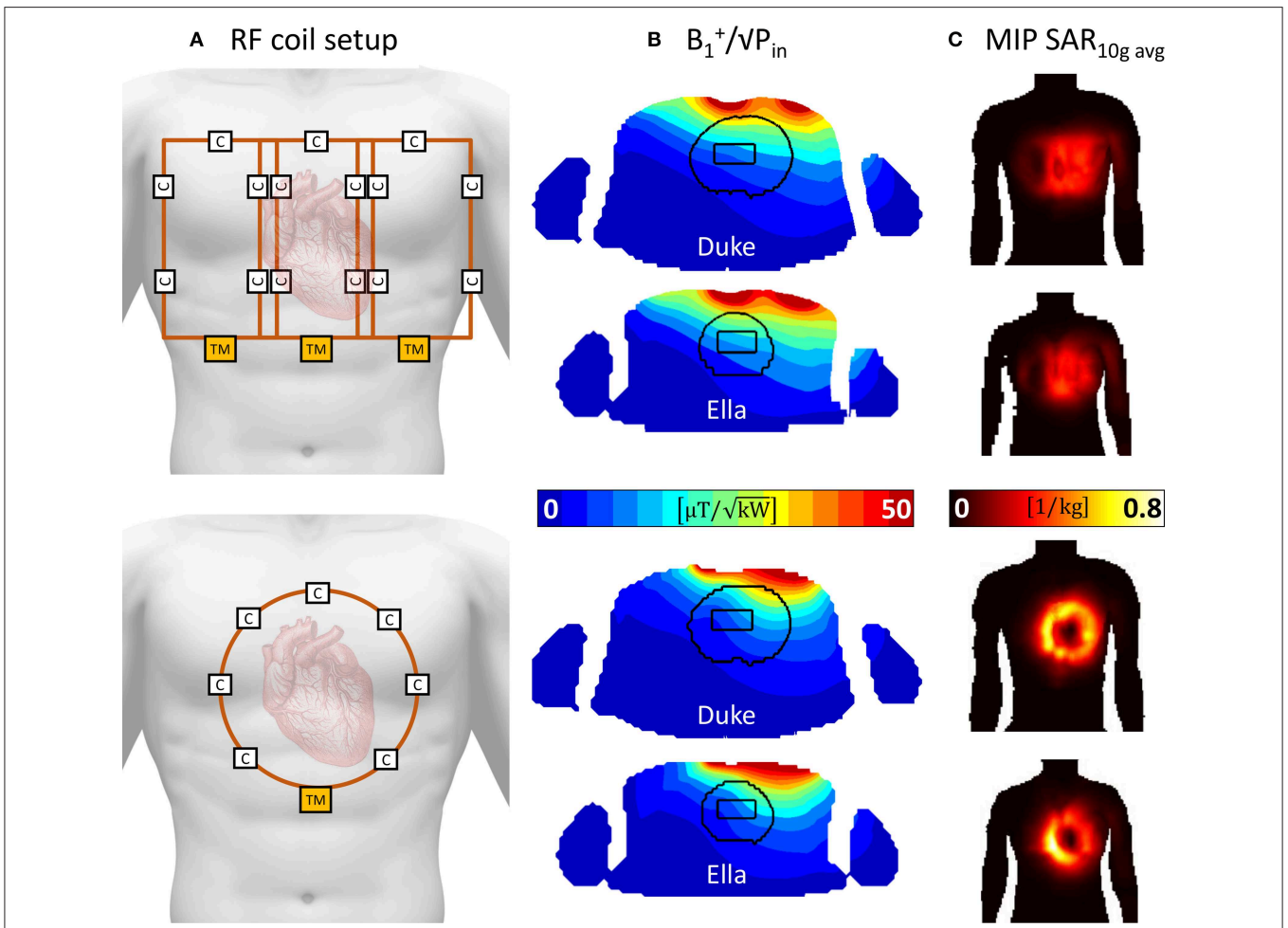


FIGURE 4 | Simulation comparison of the proposed 3-element array with a standard 14 cm single loop coil. **(A)** depicts the RF coil setups, the proposed array is depicted in the top row, whereas the single loop is depicted in the bottom part. Both coils are positioned centered over the heart. **(B)** shows the resulting $B_1^+/\sqrt{P_{in}}$ maps in a transversal slice through the center of the heart ROI. Higher B_1^+ values are seen for the 3-element arrays. **(C)** depicts coronal maximum intensity projections of the 10 g averaged specific absorption rate (SAR).

A set of 32 ^{31}P array layouts, each evaluated using two different voxel models (one male, one female) to incorporate inter-subject variability was compared via full wave 3D electromagnetic simulation with realistic loss estimations to find the best performing array design. Static B_1^+ shim phase sets optimized for a combination of

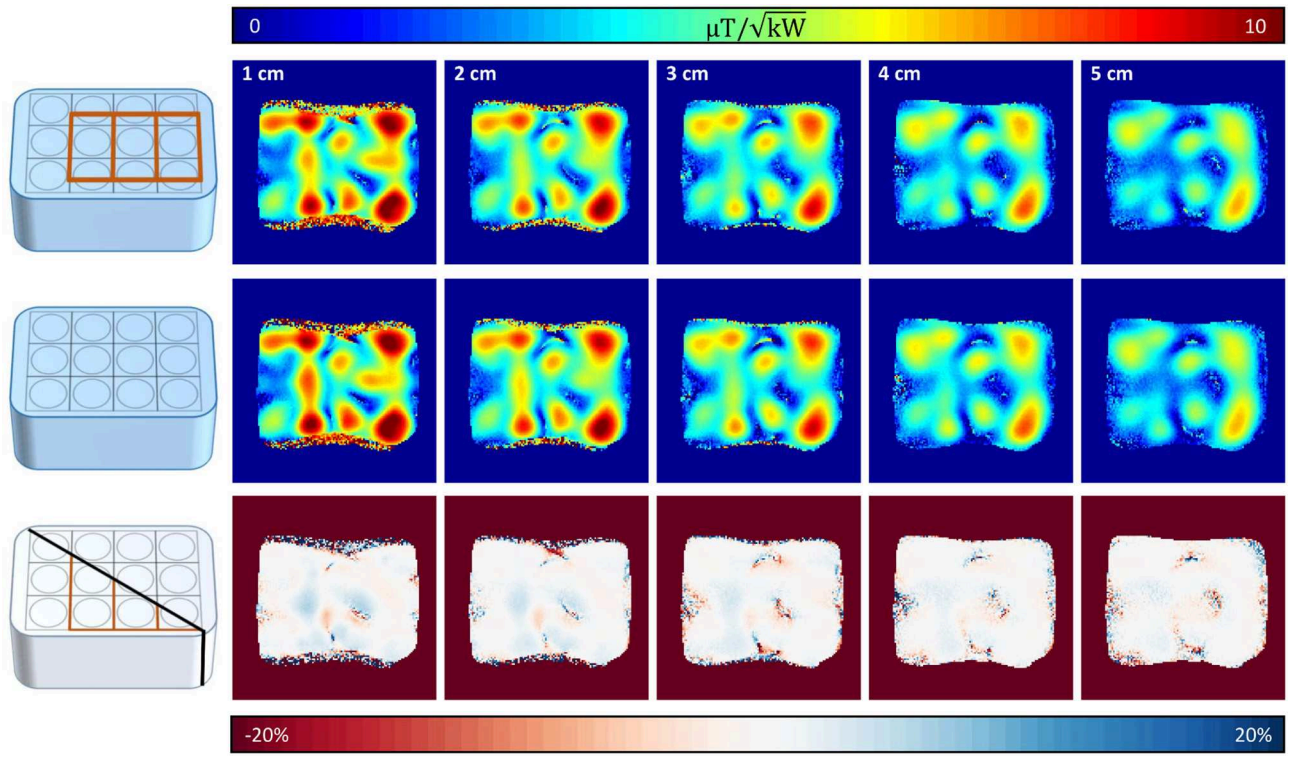


FIGURE 5 | Coronal ^{1}H B_1^+ maps with and without ^{31}P array present. Top row shows coronal B_1^+ slices (10 mm slice thickness, distance from surface/RF coil 1–5 cm) with the ^{31}P array positioned below the proton array, whereas in the middle row the same slices are shown without the ^{31}P array present. Bottom row shows the resulting difference of the B_1^+ maps above. Highest deviations are encountered where the B_1^+ mapping sequence produced artifacts due to the high B_1^+ of surface coils.

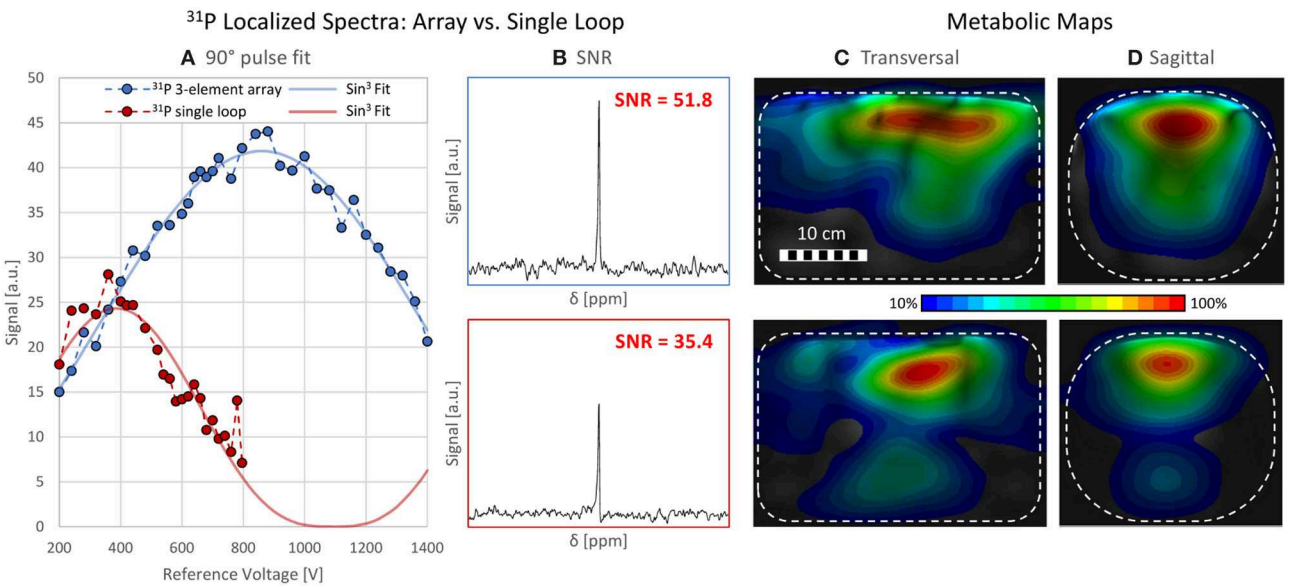


FIGURE 6 | Localized Spectroscopy and CSI data. **(A)** shows all data points acquired with a STEAM sequence from a voxel of size 50 × 20 × 50 mm^3 7 cm within a torso phantom containing ^{31}P. The blue and red data points represent the spectra acquired with the array and the loop, respectively. The course of the signal amplitude is fitted with a sin^3 function corresponding to the signal equation for STEAM sequences. The reference voltages are 400 V (single loop) and 880 V (array). **(B)** SNR comparison of localized spectra for array (SNR = 51.8, top) and the loop (SNR = 35.4, bottom). **(C,D)** Metabolic maps, interpolated from CSI data, scaled to their individual maximum, in a **(C)** transversal and **(D)** sagittal slice for the array (top) and the single loop (bottom). The phantom is depicted as a white rectangle.

homogeneity, power and SAR efficiency were calculated for all investigated arrays.

All array layouts used a fixed overlap and additional counter-wound inductances to decouple the array elements in order to save simulation time, since finding the optimal overlap for differently sized array elements is very time consuming due to the necessity to rerun the 3D simulation for each setup multiple times while changing the overlap factor slightly, until optimal decoupling is achieved. It was exemplarily shown for two elements that the differences between optimal overlap only as compared to a fixed overlap factor and the additional CWIs are negligible.

A new way of visualizing B_1^+ shimming results in the entire phase shift parameter space was introduced, allowing for quick and easy visual inspection of the variation of performance parameters on the chosen phase set. A figure of merit taking into account an equally weighted combination of homogeneity, power and SAR efficiency was employed. This approach can be universally employed for any transmit array. Depending on the requirements of the application, the weights for the figure of merit could be changed to favor a specific performance parameter. This could be useful e.g., to optimize the B_1^+ shim more strongly for SAR efficiency in applications that are SAR demanding, or for homogeneity where a uniform flip angle distribution is essential, or for power efficiency where the available transmit power is limiting.

The best performing design was a 3-element array centered above the heart with individual elements of 94 × 141 mm^2. It was integrated into the housings of the proton coil, including performance tests on the bench and in the MR scanner.

Maximum B_1^+ deviations for the ^{1}H array alone vs. the combination with the ^{31}P array were found in superficial areas and were below 20%, indicating sufficient decoupling between the two frequencies.

Simulation predicted that the proposed 3-element RF array would outperform a 14 cm single loop coil for cardiac MRS at 7 T in terms of power efficiency (+ 58% over the whole heart, + 27% in the measured VOI), SAR efficiency (+124% heart, + 81% VOI), and relative homogeneity (+30% heart, +13% VOI).

Despite the higher calculated power efficiency, in the experiment higher pulse amplitudes were necessary in the VOI for the array when compared to the single loop (**Figure 6A**). The main cause for this behavior is that the array was simulated and constructed to be optimal for human subjects, but the measurement was performed on a homogeneous phantom. Firstly, this led to a mismatch of the RF coil to the phantom load, resulting in a significant decrease of the effective input voltage at the coil ports. Secondly, the power efficiency was simulated for the array bent on a human load, but since the phantom did not allow for bending, the coil was used in flat configuration, leading to lower efficiency in depth. In addition, shielding effects from the conductive structures of the ^{1}H coil elements and interfaces were not considered in simulation and could possibly also reduce transmit efficiency. Losses associated to imperfect decoupling from the ^{1}H array, and induced common mode currents on the cable shields further contribute to the difference, although to a lesser extent, since an effort was made to keep them as small as possible.

Nevertheless, an SNR increase of + 46% in the VOI was demonstrated with identical flip angle as in the reference coil, which shows the advantages of the array in terms of receive sensitivity and supports the above reasoning for suboptimal transmit performance on the phantom.

Because of the mentioned limitations of the phantom measurement, an even stronger increase for *in vivo* measurements can be expected. In a next step, the required tests and documentation of the coil for approval of the ethics board will be established to enable the usage of the coil in an *in vivo* study.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

SR and EL designed the study. SR did the 3D simulations and drafted the manuscript. SR and MV implemented and measured the coil on the bench. SR, SW, and AS did the MR measurements and the post-processing. EL, SW, and AS revised the manuscript.

FUNDING

This work was funded by the Austrian Science Fund (FWF) grants P28059-N36 and P28867-B30.

REFERENCES

1. Roemer PB, Edelstein WA, Hayes CE, Souza SP, Mueller OM. The NMR phased array. *Magn Reson Med.* (1990) **16**:192–225. doi: 10.1002/mrm.1910160203
2. Buchli R, Meier D, Martin E, Boesiger P. Assessment of absolute metabolite concentrations in human tissue by 31P MRS *in vivo*. Part II: Muscle, liver, kidney. *Magn Reson Med.* (1994) **32**:453–8. doi: 10.1002/mrm.1910320405
3. Lee J-H, Komoroski RA, Chu W-J, Dudley JA. Methods and applications of phosphorus nmr spectroscopy *in vivo*. *Ann Rep NMR Spectr.* (2012) **75**:115–60. doi: 10.1016/B978-0-12-397018-3.00003-X
4. Bottomley PA, Wu KC, Gerstenblith G, Schulman SP, Steinberg A, Weiss RG. Reduced myocardial creatine kinase flux in human myocardial infarction an *in vivo* phosphorus magnetic resonance spectroscopy study. *Circulation.* (2009) **119**:1918–24. doi: 10.1161/CIRCULATIONAHA.108.823187
5. Yabe T, Mitsunami K, Inubushi T, Kinoshita M. Quantitative measurements of cardiac phosphorus metabolites in coronary artery disease by 31 p magnetic resonance spectroscopy. *Circulation.* (1995) **92**:15–23. doi: 10.1161/01.CIR.92.1.15
6. Neubauer S, Krahe T, Schindler R, Horn M, Hillenbrand H, Entzeroth C, et al. 31P magnetic resonance spectroscopy in dilated cardiomyopathy and coronary artery disease. Altered cardiac high-energy phosphate metabolism in heart failure. *Circulation.* (1992) **86**:1810–8. doi: 10.1161/01.CIR.86.6.1810

7. Bottomley P. Noninvasive study of high-energy phosphate metabolism in human heart by depth-resolved 31P NMR spectroscopy. *Science.* (1985) **229**:769–72. doi: 10.1126/science.4023711
8. Bottomley PA. NMR spectroscopy of the human heart. In: Harris RK, Wasylishen RL, editors. *eMagRes*. Chichester: John Wiley & Sons, Ltd. (2009). p. 1–20. doi: 10.1002/9780470034590.emrstm0345.pub2
9. Valkovič L, Clarke WT, Schmid AI, Raman B, Ellis J, Watkins H, et al. Measuring inorganic phosphate and intracellular pH in the healthy and hypertrophic cardiomyopathy hearts by *in vivo* 7T 31P-cardiovascular magnetic resonance spectroscopy. *J Cardiovasc Magn Reson.* (2019) **21**:19. doi: 10.1186/s12968-019-0529-4
10. Neubauer S, Horn M, Cramer M, Harre K, Newell JB, Peters W, et al. Myocardial Phosphocreatine-to-ATP ratio is a predictor of mortality in patients with dilated cardiomyopathy. *Circulation*. (1997) **96**:2190–6. doi: 10.1161/01.CIR.96.7.2190
11. Neubauer S. The failing heart — an engine out of fuel. *N Engl J Med.* (2007) **356**:1140–51. doi: 10.1056/NEJMra063052
12. Ladd ME, Bachert P, Meyerspeer M, Moser E, Nagel AM, Norris DG, et al. Pros and cons of ultra-high-field MRI/MRS for human application. *Prog Nucl Magn Reson Spectrosc.* (2018) **109**:1–50. doi: 10.1016/j.pnmrs.2018.06.001
13. Rodgers CT, Clarke WT, Snyder C, Vaughan JT, Neubauer S, Robson MD. Human cardiac 31P magnetic resonance spectroscopy at 7 tesla. *Magn Reson Med.* (2014) **72**:304–15. doi: 10.1002/mrm.24922
14. Hardy CJ, Bottomley PA, Rohling KW, Roemer PB. An NMR phased array for human cardiac 31P spectroscopy. *Magn Reson Med.* (1992) **28**:54–64. doi: 10.1002/mrm.1910280106
15. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Magn Reson Med.* (1999) **42**:952–62. doi: 10.1002/(SICI)1522-2594(199911)42:5<;952::AID-MRM16>;3.0.CO;2-S
16. Keil B, Wald LL. Massively parallel MRI detector arrays. *J Magn Reson.* (2013) **229**:75–89. doi: 10.1016/j.jmr.2013.02.001
17. Katscher U, Börnert P, Leussler C, van den Brink JS. Transmit SENSE. *Magn Reson Med.* (2003) **49**:144–50. doi: 10.1002/mrm.10353
18. Ullmann P, Junge S, Wick M, Seifert F, Ruhm W, Hennig J. Experimental analysis of parallel excitation using dedicated coil setups and simultaneous RF transmission on multiple channels. *Magn Reson Med.* (2005) **54**:994–1001. doi: 10.1002/mrm.20646
19. Goluch S, Kuehne A, Meyerspeer M, Kriegl R, Schmid AI, Fiedler GB, et al. A form-fitted three channel 31P, two channel 1H transceiver coil array for calf muscle studies at 7 T. *Magn Reson Med.* (2015) **73**:2376–89. doi: 10.1002/mrm.25339
20. Chmelik M, Považan M, Kršňák M, Gruber S, Tkačov M, Trattnig S, et al. *In vivo* 31 P magnetic resonance spectroscopy of the human liver at 7 T: an initial experience. *NMR Biomed.* (2014) **27**:478–85. doi: 10.1002/nbm.3084
21. Ellis J, Valkovič L, Purvis LAB, Clarke WT, Rodgers CT. Reproducibility of human cardiac phosphorus MRS (31P-MRS) at 7 T. *NMR Biomed.* (2019) **32**:e4095. doi: 10.1002/nbm.4095
22. Löring J, van der Kemp WJM, Almujayyaz S, van Oorschot JWM, Luijten PR, Klomp DWJ. Whole-body radiofrequency coil for 31P MRSI at 7T. *NMR Biomed.* (2016) **29**:709–20. doi: 10.1002/nbm.3517
23. Valkovič L, Dragonu I, Almujayyaz S, Batzakis A, Young LAJ, Purvis LAB, et al. Using a whole-body 31P birdcage transmit coil and 16-element receive array for human cardiac metabolic imaging at 7T Lundberg P, editor. *PLoS ONE.* (2017) **12**:e0187153. doi: 10.1371/journal.pone.0187153
24. Purvis LAB, Clarke WT, Valkovič L, Levick C, Pavlides M, Barnes E, et al. Phosphodiester content measured in human liver by *in vivo* 31P MR spectroscopy at 7 tesla. *Magn Reson Med.* (2017) **78**:2095–105. doi: 10.1002/mrm.26635
25. Hosseinnezhadian S, Frass-Kriegl R, Goluch-Roat S, Pichler M, Sieg J, Vít M, et al. A flexible 12-channel transceiver array of transmission line resonators for 7 T MRI. *J Magn Reson.* (2018) **296**:47–59. doi: 10.1016/j.jmr.2018.08.013
26. Betts JG, Desaix P, Johnson EW, Johnson JE, Korol O, Kruse D, et al. *Anatomy and Physiology*. Houston, TX : OpenStax College, Rice University. (2017).
27. Kozlov M, Turner R. Analysis of RF transmit performance for a 7T dual row multichannel MRI loop array. Conf Proc IEEE Eng Med Biol Soc. (2011) **2011**:547–53. doi: 10.1109/IEMBS.2011.6090101
28. Mispelter J, Lupu M, Briguet A. *Nmr Probeheads for Biophysical and Biomedical Experiments: Theoretical Principles and Practical Guidelines*. 1st ed. London: Imperial College Press. (2006). doi: 10.1142/p438
29. Lee RF, Giaquinto RO, Hardy CJ. Coupling and decoupling theory and its application to the MRI phased array. *Magn Reson Med.* (2002) **48**:203–13. doi: 10.1002/mrm.10186
30. Kumar A, Edelstein WA, Bottomley PA. Noise figure limits for circular loop MR coils. *Magn Reson Med.* (2009) **61**:1201–9. doi: 10.1002/mrm.21948
31. Wright SM, Wald LL. Theory and application of array coils in MR spectroscopy. *NMR Biomed*. (1997) **10**:394–410. doi: 10.1002/(SICI)1099-1492(199712)10:8<;394::AID-NBM494>;3.0.CO;2-0
32. Meyerspeer M, Serés Roig E, Gruetter R, Magill AW. An improved trap design for decoupling multinuclear RF coils. *Magn Reson Med.* (2014) **72**:584–90. doi: 10.1002/mrm.24931
33. Wilcox M, McDougall M. Double-Tuned cable traps for multinuclear MRI And MRS. In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Honolulu, HI (2018). p. 1364–7. doi: 10.1109/EMBC.2018.8512474
34. Seeber DA, Jevtic J, Menon A. Floating shield current suppression trap. *Concepts Magn Reson.* (2004) 21B:26–31. doi: 10.1002/cmr.b.20008
35. Robson MD, Tyler DJ, Neubauer S. Ultrashort TE chemical shift imaging (UTE-CSI). *Magn Reson Med.* (2005) **53**:267–74. doi: 10.1002/mrm.20344
36. Frahm J, Merboldt KD, Hänicke W. Localized proton spectroscopy using stimulated echoes. *J Magn Reson.* (1987) **72**:502–8. doi: 10.1016/0022-2364(87)90154-5
37. Purvis LAB, Clarke WT, Biasiolli L, Valkovič L, Robson MD, Rodgers CT. OXSA: an open-source magnetic resonance spectroscopy analysis toolbox in MATLAB Motta A, editor. *PLoS ONE.* (2017) **12**:e0185356. doi: 10.1371/journal.pone.0185356
38. Vanhamme L, van den Boogaart A, Van Huffel S. Improved method for accurate and efficient quantification of MRS data with use of prior knowledge. *J Magn Reson.* (1997) **129**:35–43. doi: 10.1006/jmre.1997.1244
39. Rodgers CT, Robson MD. Receive array magnetic resonance spectroscopy: whitened singular value decomposition (WSVD) gives optimal bayesian solution. *Magn Reson Med.* (2010) **63**:881–91. doi: 10.1002/mrm.22230
40. Chung S, Kim D, Breton E, Axel L. Rapid B1+ mapping using a preconditioning RF pulse with TurboFLASH readout. *Magn Reson Med.* (2010) **64**:439–46. doi: 10.1002/mrm.22423

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Maria Piarulli is, since September 2018, an Assistant Professor at the Washington University in St. Louis (WUSTL), in USA. She arrived at WUSTL after two years as post-doctoral fellowship at the Argonne National Laboratory (ANL), which has been her first and only post-doctoral fellowship after her graduation at the Old Dominion University (ODU) of Norfolk (August 2015). Since her graduate studies, Maria has been involved in outstanding research projects, ranging from the derivation of a consistent model for the nuclear interactions and currents withing chiral effective field theory, to the application of ab-initio techniques to study the nuclear structure of light nuclei up to 12C and nuclear matter. Her work has received great recognition, as demonstrated by the Outstanding PhD Dissertation Award for 2015-2016 that she has received from the ODU College of Sciences, and by the many plenary talk invitations to international conferences, as, for instance, the International Nuclear Physics Conference - INPC2019 (Glasgow, UK, July 2019).

In this paper, the authors present ab-initio calculations of the reduced matrix elements entering β-decays and electron captures in light nuclei with $A \leq 10$, using potentials and consistent weak axial currents, developed in chiral effective field theory, and the Quantum Monte Carlo method to calculate the nuclear wave functions. The agreement with the experimental data is quite nice, with a two-body axial current contribution usually of the order of 3%. The only exception is represented by the $A = 8$ systems, for which the reduced matrix elements are severely underpredicted by the theory, even after the inclusion of the (large, about 30 %) contribution from two-body axial currents. This severe underprediction indicates the need of further improvements in the $A = 8$ nuclear wave functions. In conclusion, with this work, the authors have set the foundations for the development of an accurate and unified understanding of weak processes. In a near future the neutrino-nucleus interactions will be at reach, providing crucial inputs for long-base neutrino oscillation experiments like, for instance, MiniBooNE, T2K, Minerva and the upcoming DUNE.

Miha Mihovilovic graduated from the Jozef Stefan Institute of Ljubljana (Slovenia) in 2012, with a thesis on the measurement of double-polarized asymmetries in quasi-elastic electron scattering on 3He, an experiment conducted within the Hall A collaboration at the Jefferson Laboratory (JLab), in USA. During his graduate studies, he also participated in 16 different experiments conducted at JLab. He was then a post-doctoral fellow at the Institute for Nuclear Physics in Mainz (Germany), where he was in charge of a project dedicated to the study of electromagnetic form factors at extremely small momentum transfer. In 2016 he returned in Slovenia with his second post-doctoral fellowship, after which he has become shared research associate of the University of Mainz, the University of Ljubljana, and the Jo˘zef Stefan Institute. He is involved in many projects, among which those dedicated to discovering hyper-nuclei and to finding the dark photon. His research activity has received great recognition, as demonstrated by his qualification in a global competition among young scientists worldwide to participate in the 65th Lindau Nobel Laureate Meeting, and by the many plenary talk invitations to international conferences, as, for instance, the XXII International Conference on Few-Body Problems in Physics (FB22), in 2018. He is currently a member of the A1 collaboration at the University of Mainz, the Hall A collaboration at JLab, and the NUSTAR collaboration in Darmstadt.

The "proton radius puzzle" is one of the most intriguing problems in physics since a decade or so. The proton (charge) radius has been measured within different techniques, i.e. hydrogen spectroscopy, elastic lepton (electron) scattering, and muonic hydrogen spectroscopy. The results from these different techniques were found back in 2010 significantly different, with the muonic hydrogen spectroscopy result about 4 % smaller than the results obtained with the other techniques. This puzzle clearly has motivated several new experiments and different reanalyses of the existing data, and several steps have been done to find the solution. In this paper, the authors revisit the first electron scattering data published back in 1963 and used in the standard dipole parametrization of the proton form factor. In the reanalysis, they have discovered a sign error in the original work which would have led to a value for the radius in good

agreement with the latest muonic hydrogen spectroscopy result. The authors additionally performed a new analysis of the data, based on a Monte Carlo study of different form factor models, a tool not available in the 1960s. Within this reanalysis, the authors give a more reliable determination of the radius, which is found in very good agreement with recent extractions of the radius from other techniques and with the new recommended value. In conclusion, the "proton radius puzzle" seems to be nowadays essentially solved.

frontiers in Physics

ORIGINAL RESEARCH
published: 02 November 2020
doi: 10.3389/fphy.2020.00363

Weak Transitions in Light Nuclei

Garrett B. King[1], *Lorenzo Andreoli*[1], *Saori Pastore*[1,2] *and Maria Piarulli*[1]*

[1] *Department of Physics, Washington University, St. Louis, MO, United States,* [2] *McDonnell Center for the Space Sciences at Washington University, St. Louis, MO, United States*

Nuclei are used for high-precision tests of the Standard Model and for studies of physics beyond the Standard Model. Without a thorough understanding of nuclei, we will not be able to meaningfully interpret the growing body of experimental data nor will we be able to disentangle new physics signals from underlying nuclear effects. This calls for accurate calculations of nuclear structure and reactions. In this work, we focus on electroweak decays in nuclei with mass number $A \leq 10$ and report on *ab initio* Quantum Monte Carlo calculations of reduced matrix elements entering beta decays and electron captures in nuclei with mass number $A \leq 10$. The many-body wave functions are calculated using selected Norfolk two- and three-nucleon potential models and associated one- and two-body axial currents at tree-level obtained from a chiral effective field theory with pions, nucleons, and Δ. The agreement with the experimental data is satisfactory except for transitions in $A = 8$ nuclei. In this specific case, the theory significantly underpredicts the experimental data, which indicates the need of further improvements in the corresponding nuclear wave functions. In this study, emphasis is placed on the contributions of two-body axial currents that are carefully analyzed using two-body transition densities. This allow us to study the spatial distribution and short-range behavior of two-body dynamics. In particular, the transition densities when scaled to peak at 1.0 exhibit universal short-range behavior across the considered nuclei, while they differ in the long-range tails.

Keywords: nuclear interactions, nuclear currents, chiral effective field theory, *ab-initio* calculations, weak transitions

OPEN ACCESS

Edited by:
Nunzio Itaco,
University of Campania Luigi Vanvitelli, Italy

Reviewed by:
Maria Colonna,
Laboratori Nazionali del Sud (INFN), Italy
Kevin Fossez,
Michigan State University, United States

***Correspondence:**
Maria Piarulli
mpiarulli@physics.wustl.edu

Specialty section:
This article was submitted to Nuclear Physics, a section of the journal Frontiers in Physics

Received: *08 June 2020*
Accepted: *28 June 2020*
Published: *02 November 2020*

Citation:
King GB, Andreoli L, Pastore S and Piarulli M (2020) Weak Transitions in Light Nuclei. Front. Phys. 8:363. doi: 10.3389/fphy.2020.00363

1. INTRODUCTION

Nuclei are used for high-precision tests of the Standard Model and for studies of physics beyond the Standard Model. Without a thorough understanding of nuclei, we will not be able to meaningfully interpret the growing body of experimental data nor will be able to disentangle new physics signals from underlying nuclear effects. Current and next generation experimental programs are poised to address open questions within fundamental symmetries and neutrino physics to understand the origin of nonzero neutrino masses, the observed matter, and anti-matter unbalance, and the nature of dark matter. These experimental endeavors often rely on accurate calculations of electroweak structure and reactions in nuclei.

For example, nuclear physics plays a pivotal role in searches for neutrinoless double beta ($0\nu\beta\beta$) decays, which are the subject of an intense experimental research program [1–14]. In these decays, two neutrons inside the nucleus decay into two protons via the exchange of a neutrino, emitting two electrons. The rates of these decays depend not only on unknown fundamental neutrino parameters but also on nuclear properties. Extracting the neutrino parameters from experiments requires theoretical evaluation of nuclear matrix elements for neutrinoless double beta decay. This

decay, if observed, would have tremendous theoretical implications and could give insight into our understanding of the observed matter-antimatter asymmetry in the universe. Calculations for nuclei of experimental interest ($A \geq 48$) are based on computational methods that inevitably adopt approximations to solve the nuclear many-body problem—e.g., model space truncations and/or omission of many-body effects. As a consequence, estimates of $0\nu\beta\beta$ matrix elements may vary by a factor of two when computed using different computational models (see [15] and references therein). It is then crucial to understand neutrino–nucleus interactions with great accuracy as well as the role and relevance of many-body dynamics, such as many-nucleon correlations and currents.

Furthermore, intense experimental research activity is currently focused on long-base neutrino oscillation experiments (such as MiniBooNE, T2K, MicroBooNE, Minerνa, and the upcoming DUNE; [16–20]) aimed at profiling neutrinos, whose masses, among other properties, are still not known. Neutrinos signal their presence by interacting with nuclei which are the active material in the detectors. Additionally, in this case, meaningful interpretations of the data require an accurate understanding of the way neutrinos interact with nuclei.

The study of light nuclei, for which the nuclear many-body problem can be solved exactly or within controlled approximations by fully retaining many-nucleon correlations and electroweak currents, offers the possibility of quantifying the contribution from many-body effects and consequently of assessing the robustness of a given approximation. In this work, we report on a recent study of electroweak matrix elements in $A \leq 10$ nuclei entering single beta decays and electron captures (or inverse beta decays). Rates of single beta decay—a process in which a proton (neutron) inside the nucleus decays into a neutron (proton) with the emission of a positron (electron) and an electron (anti)neutrino—are, in most cases, experimentally well-known. This provides us with stringent means to validate our theoretical description of nuclear systems and to assess the role of many-body dynamics. In particular, we work within the nuclear microscopic approach in which nuclei are described in terms of non-relativistic nucleons interacting with each other via two- and three-body nuclear potentials and with external probes, such as neutrinos, electrons, and photons, via one- and two-body current operators. We use Quantum Monte Carlo (QMC) computational methods [21–23] to solve the many-body nuclear problem with a nuclear Hamiltonian consisting of high-quality two- and three-body potentials obtained from a chiral effective field theory (χEFT) that retains nucleons, pions, and Δ-isobars as explicit degrees of freedom [24–28]. We base the calculation of the transition matrix elements on one- and two-body axial currents [29] and provide results for one- and two-body weak transition densities. The latter will turn out to be particularly important to understand the role of short-range many-body dynamics.

Ab initio studies in light nuclei allow us to carefully test many-body correlations and electroweak currents and serve as benchmark to approximated many-body methods currently employed to access heavier nuclear systems [30, 31]. A study along these lines has been carried out recently in [32].

This study represents a first step into the validation of our theoretical model. In fact, beta decay processes occur at zero momentum transfer while the energy transfer involved is of the order of a few MeVs. Neutrinos exchanged in $0\nu\beta\beta$ processes carry a value of momentum transfer of the order of few hundreds of MeV/c [15], while the energy transfer in neutrino oscillation experiments covers a large phase space reaching the GeV scale. It is then essential to validate the theoretical model in a wide range of energy and momentum transfer to have a complete and unified description of neutrino–nucleus interactions. For example, calculations of total and partial muon-capture rates and comparisons with the known experimental data will probe our model at intermediate values of momentum transfer and will be the subject of our future work. At higher energies, neutrino–nucleus cross sections calculations [33, 34] are the main input to interpret the data from long and short baseline neutrino-oscillation experiments, which use nuclei as active material in the detectors. Specifically, current challenges concern the implementation of microscopic models of nuclear dynamics—that fully capture correlation effects—in neutrino event generators [35] used to simulate the neutrino interaction physics.

The paper is structured as follows: In section 2, we briefly summarize the theoretical and computational methods adopted in the present work and refer the interested reader to [36] for further details. In sections 3 and 4, we present our results and conclusions.

2. THEORY

2.1. Quantum Monte Carlo Methods

Quantum Monte Carlo methods have been most recently described in several review articles [21–23]. Here, we briefly sketch the employed calculational scheme and refer the reader to [36] for details. We use both Variational Monte Carlo (VMC) and Green's Function Monte Carlo (GFMC) methods to calculate transitions matrix elements. We base our study on the following many-body Hamiltonian:

$$H = \sum_i K_i + \sum_{i<j} v_{ij} + \sum_{i<j<k} V_{ijk} \tag{1}$$

where K_i is the non-relativistic kinetic energy operator and v_{ij} and V_{ijk} are the NV2 and NV3 Norfolk local chiral interactions developed in [24–28]. Together, we denote these interactions as NV2+3.

For a nuclear state with given angular momentum and parity J^π, isospin T, and isospin projection T_Z, the VMC method takes as a starting point a trial wave function $\Psi_V(J^\pi, T, T_z)$, constructed as follows

$$|\Psi_V\rangle = \mathcal{S}\prod_{i<j}^{A}\left[1 + U_{ij} + \sum_{k\neq i,j}^{A}\tilde{U}_{ijk}^{TNI}\right]|\Psi_J\rangle. \tag{2}$$

The Jastrow wave function Ψ_J is fully antisymmetric and has the $(J^\pi; T, T_z)$ quantum numbers of the state of interest, and U_{ij} and

$\tilde{U}_{ijk}^{TNI}$ are two- and three-body correlation operators that reflect the influence of the two- and three-body forces, respectively [37–40]. The state $|\Psi_V\rangle$ has embedded variational parameters that one adjusts to minimize the expectation value

$$E_V = \frac{\langle\Psi_V|H|\Psi_V\rangle}{\langle\Psi_V|\Psi_V\rangle} \geq E_0 \tag{3}$$

evaluated with Metropolis Monte Carlo integration [41].

The VMC wave function $|\Psi_V\rangle$ is further improved using the GFMC method. The variational state is propagated in imaginary time with the operator $\exp\left[-(H-E_0)\tau\right]$. This is done in small steps in imaginary time, $\Delta\tau$, and produces the following:

$$\Psi(\tau) = e^{-(H-E_0)\tau}\Psi_V = \left[e^{-(H-E_0)\Delta\tau}\right]^n \Psi_V \tag{4}$$

One can see that, for $\tau \rightarrow \infty$, the variational state becomes the desired state Ψ_0. The evaluation of the off-diagonal expectation value of a given operator O is calculated using the following approximation

$$\begin{aligned}&\frac{\langle\Psi^f(\tau)|O|\Psi^i(\tau)\rangle}{\sqrt{\langle\Psi^f(\tau)|\Psi^f(\tau)\rangle}\sqrt{\langle\Psi^i(\tau)|\Psi^i(\tau)\rangle}}\\ &\approx \langle O(\tau)\rangle_{M_i} + \langle O(\tau)\rangle_{M_f} - \langle O\rangle_V,\end{aligned} \tag{5}$$

where $\langle O\rangle_V$ is the variational expectation value and $\langle O(\tau)\rangle_M$ is the mixed estimate defined as the following:

$$\langle O(\tau)\rangle_{M_f} = \frac{\langle\Psi^f(\tau)|O|\Psi_V^i\rangle}{\langle\Psi^f(\tau)|\Psi_V^f\rangle}\sqrt{\frac{\langle\Psi_V^f|\Psi_V^f\rangle}{\langle\Psi_V^i|\Psi_V^i\rangle}}, \tag{6}$$

and $\langle O(\tau)\rangle_{M_i}$ is defined similarly (see [42] for more details).

2.2. Norfolk Interaction Models

The calculations of weak transitions presented in this work employ the high-quality local NV2+3 interactions developed in [24–27]. The two-nucleon potentials, NV2s, include a strong interaction component derived from a χEFT that involves nucleons, pions, and Δ-isobars as explicit degrees of freedom and an electromagnetic interaction component, including up to terms quadratic in the fine structure constant α. The component induced by the strong interaction is separated into long- and short-range parts, labeled v_{ij}^L and v_{ij}^S, respectively. The v_{ij}^L part is mediated via one-pion-exchange (OPE) and two-pion-exchange (TPE) terms up to next-to-next-to-leading order (N2LO) in the chiral expansion. Its strength is determined by the nucleon axial coupling g_A and the nucleon-to-Δ axial coupling h_A, the pion decay amplitude F_π, and LECs c_1, c_2, c_3, c_4, and $b_3 + b_8$ constrained by fits to πN scattering-data [43]. Values of these LECs are provided in Table 1 of [36].

The pion-range operators are strongly singular at short-range in configuration space and are regularized by a radial function that is characterized by a cutoff R_L as reported in [24–27]. The v_{ij}^S part, however, is described by contact terms up to next-to-next-to-next-to-leading order (N3LO), characterized by 26 unknown LECs. These interactions have been recently constrained to a large set of *NN*-scattering data, as assembled by the Granada group [44–46], including the deuteron ground-state energy and two-neutron scattering length. For the contact terms, we use a Gaussian representation of the three-dimensional delta function, with R_S being the short-range regulator.

In this work, we focus on one class of NV2 interaction, namely the NV2-Ia. This class fits about 2,700 *NN* scattering data in the range of 0–125 MeV of laboratory energies with a χ^2/datum $\lesssim 1.1$ [24, 25]. The NV2-Ia uses the combination of short- and long-range regulators $(R_S, R_L) = (0.8, 1.2)$ fm.

The NV2 models alone are not enough to provide sufficient attraction in GFMC calculations of the binding energies of light nuclei [25]. For this reason, a consistent three-body interaction up to N2LO in the chiral expansion has been developed [47] to go with the two-body potential. This interaction consists of a long-range part mediated by two-pion exchange and a short-range part parameterized in terms of two contact interactions [48, 49]. The two 3*N* LECs, namely c_D and c_E, have been obtained either by fitting exclusively strong-interaction observables [47, 50–52] or by relying on a combination of strong- and weak-interaction ones [27, 53, 54]. This last approach is made possible by a relation established in χEFT [55] between c_D and the LECs entering the contact axial current at N3LO [53, 54], [Schiavilla, private communication].

In [47], the values for c_D and c_E were obtained by reproducing both the experimental trinucleon ground-state energies and *nd* doublet scattering length for each of the NV2 models considered. On the other hand, in [27], these LECs were constrained by fitting, in addition to the trinucleon energies, the empirical value of the Gamow-Teller matrix element in tritium β-decay. The resulting Hamiltonian is denoted as NV2+3-Ia (or Ia for short) in the first case, and as NV2+3-Ia* (or Ia* for short) in the second.

As shown in **Table 1**, these two different procedures for fixing c_D and c_E produced rather different values for these LECs, particularly for c_E which was found to be relatively large and negative in the unstarred models but quite small, and not consistently negative, in the starred models. This in turn impacts predictions for the spectra of light nuclei [36] and the equation of state of neutron matter, since a negative c_E leads to repulsion in light nuclei but attraction in neutron matter [56].

The starred and unstarred NV2+3 Norfolk interactions have been implemented in both the VMC and GFMC codes and used to perform calculations of the energy levels [28, 47], charge radii, and longitudinal elastic form factors [23] of $A = 4 - 12$ nuclei that are found to be in very satisfactory agreement with the experimental data. Furthermore, two of the NV2+3* models have been also used to perform VMC calculations of the Fermi, Gamow-Teller, and tensor densities for $^6\text{He} \rightarrow {}^6\text{Be}$ and $^{12}\text{Be} \longrightarrow {}^{12}\text{C}$ transitions [57], relevant for studies of $0\nu\beta\beta$.

The NV2 models have recently been used in benchmark calculations of the energy per particle of pure neutron matter as a function of the baryon density using three independent many-body methods: Brueckner-Bethe-Goldstone (BBG), Fermi hypernetted chain/single-operator chain (FHNC/SOC), and AFDMC [58]. The inclusion of three-body forces is essential for a realistic description of neutron matter. These types of calculation

TABLE 1 | c_D and c_E values of the contact terms in the three-nucleon interactions obtained from fits to (i) the *nd* scattering length and the trinucleon binding energies [27, 47]; and (ii) the central value of the ^{3}H GT matrix element and the trinucleon binding energies (starred values).

	Ia	Ia*
c_D	3.666	−0.635
c_E	−1.638	−0.090

See text and [36] for details.

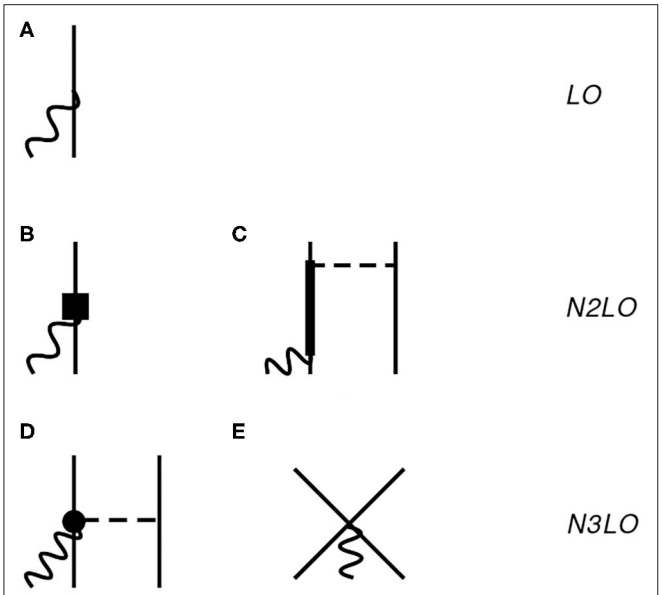


FIGURE 1 | Diagrams illustrating the contributions to the axial current up to N3LO (Q^0) used in this work. Nucleons, Δ-isobars, pions, and external fields are denoted by solid, thick-solid, dashed, and wavy lines, respectively. The diagrams in **(A,B)** represent the leading order one-body term and its relativistic correction (denoted by the square vertex), respectively. The remaining diagrams, **(C–E)** provide two-body corrections to the one-body terms. The dot in **(D)** denotes a vertex induced by subleading terms in the π-nucleon chiral Lagrangian [26]. Only a single time ordering is shown.

are particularly relevant for the quantitative assessment of the systematic error of the different many-body approaches and how they depend upon the nuclear interaction of choice.

Preliminary AFDMC calculations of the equation of state of pure neutron matter carried out with the unstarred NV2+3 Norfolk interactions [Piarulli et al., private communication] are not compatible with the existence of two solar masses neutron stars, in conflict with recent observations [59, 60]. However, the smaller values of c_E in the 3N force of the starred NV2+3 potentials might mitigate, if not resolve this problem, while predicting light-nuclei spectra <4% away from the experimental data [36]. Studies along these lines are under way.

2.3. Axial Currents in χEFT

Many-body currents are crucial for providing a quantitatively successful description of many nuclear electroweak observables [61], such as nuclear electromagnetic form factors [62–65], low-energy electroweak transitions [66–73], and electroweak scattering [33]. They have also been used in studies of double beta decay matrix elements [32, 57, 74, 75]. The study of the electroweak matrix elements carried out in this work employs one- and two-body axial currents derived within the same χEFT used for the NV2+3 interactions [27]. We use two-body axial currents at tree-level constructed in [26, 27, 29]. Here, we briefly describe the contributions shown in **Figure 1** and refer the reader to [27, 36] for the explicit expressions of the current operators and for the tables reporting the values of the parameters adopted in the present work. We note that χEFT electroweak operators have been also derived in [76–79].

In **Figure 1A**, the LO term, which scales as Q^{-3} in the power counting (Q denotes generically a low-momentum scale), is given by the standard Gamow-Teller one-body operator

$$\mathbf{j}^{\rm LO}_{5,a}(\mathbf{q}) = -\frac{g_A}{2}\tau_{i,a}\boldsymbol{\sigma}_i e^{i\mathbf{q}\cdot\mathbf{r}_i}, \tag{7}$$

where g_A is the nucleon axial coupling constant ($g_A = 1.2723$ [80]) $\boldsymbol{\sigma}_i$ and $\boldsymbol{\tau}_i$ are the spin and isospin Pauli matrices of nucleon i, $\mathbf{q}$ is the external field momentum, $\mathbf{r}_i$ is the spacial coordinate of nucleon i, and $a = x, y, z$.

We account for an additional one body-operator shown in **Figure 1B**. This contribution enters at N2LO (or Q^{-1}) in the chiral expansion and represents a relativistic correction to the single-nucleon operator at LO. At N2LO there is the appearance of the leading two-body contribution illustrated in **Figure 1C** by the tree-level diagram involving the excitation of a Δ-isobar. In the tables and figures below, we label these two contributions as N2LO-RC and N2LO-Δ, respectively. Following the same notation introduced in [27], we write the cumulative N2LO contribution as

$$\mathbf{j}^{\rm N2LO}_{5,a}(\mathbf{q}) = \mathbf{j}^{\rm N2LO}_{5,a}(\mathbf{q};\,{\rm RC}) + \mathbf{j}^{\rm N2LO}_{5,a}(\mathbf{q};\,\Delta). \tag{8}$$

Finally, the N3LO contributions (scaling as Q^0) involve a term of one-pion range illustrated in **Figure 1D** and a contact term shown in **Figure 1E**, which together give the following N3LO correction

$$\mathbf{j}^{\rm N3LO}_{5,a}(\mathbf{q}) = \mathbf{j}^{\rm N3LO}_{5,a}(\mathbf{q};\,{\rm OPE}) + \mathbf{j}^{\rm N3LO}_{5,a}(\mathbf{q};\,{\rm CT}). \tag{9}$$

These terms are denoted with N3LO-OPE and N3LO-CT, respectively, and their expression are reported in Equations (2.7)–(2.10) of [27]. As discussed at length in [36], the N3LO-CT contact current involves the LEC c_D, which also enters the three-nucleon force. The values for the LEC c_D used in this work are reported in **Table 1** and are changed consistently in the axial current depending on the nuclear interaction used to construct the wave functions. In this work, we especially focus on the nuclear interactions NV2+3-Ia and NV2+3-Ia*.

TABLE 2 | Gamow-Teller RMEs in A = 6, 7, 8, and 10 nuclei obtained with chiral axial currents [27] and VMC wave functions corresponding to the NV2+3-Ia/Ia* Hamiltonian models [24, 25, 27, 47].

Transition	Model	LO	N2LO-RC	N2LO-Δ	N3LO-OPE	N3LO-CT	(Total-LO)	Total	Expt.
^{6}He($0^+;1$)→ ^{6}Li($1^+;0$)	Ia	2.200	−0.016	0.037	0.039	−0.005	0.056	2.256	2.1609 (40)
[42]→[42]	Ia*	2.192	−0.015	0.036	0.038	−0.054	0.005	2.197	
^{7}Be($\frac{3}{2}^-;\frac{1}{2}$)→^{7}Li($\frac{3}{2}^-;\frac{1}{2}$)	Ia	2.317	−0.024	0.099	0.083	−0.010	0.148	2.465	2.3556 (47)
[43]→[43]	Ia*	2.327	−0.024	0.098	0.082	−0.121	0.036	2.362	
^{7}Be($\frac{3}{2}^-;\frac{3}{2}$) →^{7}Li($\frac{1}{2}^-;\frac{1}{2}$)	Ia	2.157	0.000	0.066	0.063	−0.009	0.121	2.278	2.1116 (57)
[43]→[43]	Ia*	2.158	0.000	0.065	0.063	−0.103	0.026	2.184	
^{8}Li($2^+;1$)→^{8}Be($2^+;0$)	Ia	0.147	0.000	0.032	0.011	−0.001	0.041	0.188	0.284 [84]
[431]→[44]	Ia*	0.141	0.000	0.031	0.010	−0.017	0.025	0.166	0.190 [85]
^{8}B($2^+;1$)→^{8}Be($2^+;0$)	Ia	0.146	0.000	0.032	0.011	−0.001	0.042	0.188	0.269 (20)
[431]→[44]	Ia*	0.148	0.000	0.032	0.010	−0.016	0.026	0.174	
^{8}He($0^+;2$)→^{8}Li($1^+;1$)	Ia	0.386	−0.004	0.034	0.009	−0.001	0.038	0.424	0.512 (6)
[422]→[431]	Ia*	0.362	−0.004	0.035	0.009	−0.010	0.029	0.391	
^{10}C($0^+;1$)→^{10}B($1^+;0$)	Ia	1.940	−0.024	0.026	0.042	−0.006	0.039	1.9879	1.8331 (34)
[442]→[442]	Ia*	2.051	−0.012	0.020	0.039	−0.065	−0.017	2.033	

Columns labeled with "LO," "N2LO-RC," "N2LO-Δ," "N3LO-OPE," and "N3LO-CT" refer to the contributions given by the diagrams illustrated in ***Figures 1A–E****, respectively. The cumulative results are reported in the column labeled "Total." Experimental values from [81–85] are given in the last column. The dominant spatial symmetry of the VMC wave function are reported in the first column. Statistical errors associated with the Monte Carlo integrations are not shown but are below 1%. Values for the N2LO-RC for transitions to states in ^{8}Be and to the state ^{7}Li($\frac{3}{2}^-;\frac{1}{2}$) are 0.000 within the statistical uncertainty of the integration.*

3. RESULTS

3.1. VMC Reduced Matrix Elements

In this section, we present the results of calculations of GT reduced matrix elements (RMEs), defined as the following:

$$\text{RME} = \frac{\sqrt{2J_f+1}}{g_A}\frac{\langle J_f M|j^z_{5,\pm}|J_i M\rangle}{\langle J_i M, 10|J_f M\rangle} \tag{10}$$

where $j^z_{5,\pm}$ is the z-component in the limit $\mathbf{q}\to 0$ of the charge-raising/lowering current $\mathbf{j}_{5,\pm}= \mathbf{j}_{5,x} \pm i\mathbf{j}_{5,y}$, and $\langle J_i M, 10|J_f M\rangle$ is a Clebsch-Gordan coefficient. **Table 2** summarizes the results of the VMC calculation of GT RMEs. These calculations were evaluated with variational wave functions generated using the NV2+3-Ia and NV2+3-Ia* nuclear Hamiltonians. The table breaks the calculations down order-by-order: the LO contribution from the one-body axial current in **Figure 1A**, the N2LO contributions coming from a one-body relativistic correction to the LO term (**Figure 1B**) and a two-body contribution involving the excitation of a nucleon into a Δ by pion exchange (**Figure 1C**), and the contributions at N3LO from the one pion exchange (**Figure 1D**) and the contact term (**Figure 1E**). The sum of the two body contributions (Total-LO) and the total RME are given in addition to the breakdown of each contribution. The dominant spatial symmetries [86] of the wave functions in each calculation are listed below the transition in **Table 2**. Experimental values from [81–85] are listed in the last column of **Table 2**.

From these results, we see that in the A = 3, 6, 7, and 10 cases, the LO contribution is ≈ 97% of the total RME in VMC calculations. The other ≃ 3% are made up of beyond leading order contributions. While beyond leading order contributions only make up a small percentage of the total RME for the A = 3, 6, 7, and 10 cases, they are a much larger contribution for A = 8, making up ∼ 20–30% of the total RME. This is attributed to a difference in the dominant spatial symmetries between the initial and final states of the A = 8 VMC wave functions, resulting in a smaller overlap between the initial and final wave functions and a consequent suppression of the GT RME at leading order. This indicates that the wave functions are lacking correlations and that an improvement of the theoretical prediction will require further developments of the wave functions, such as the inclusion of more correlations and development of better constrained small components. As similar behavior is also found in the calculations of [87]. The total two-body contribution is typically an enhancement of the total RME, except in the A = 10 case for model NV2+3-Ia*, where the matrix element is reduced. The short-range behavior of the two-body corrections is studied in detail in section 3.2 where we analyze the two-body transition densities.

For ^{8}Li →^{8} Be beta decay, there are two different log(ft) values in the literature that provide very different values for the GT matrix element. In **Table 2**, we present the RMEs for this decay using the ft values from [84, 85], obtained with the following formula [83]:

$$\text{RME(EXPT)} = \frac{1}{g_A}\sqrt{2J_i+1}\sqrt{\frac{6139\pm 7}{ft}}, \tag{11}$$

where J_i is the angular momentum of the initial state. Note that the Fermi transition strength is negligible in deriving Equation (11). This formula uses the value g_A = 1.2723. Additionally, we note that in [82, 88], a value of 6,147 is used in place of 6,139. Even with this uncertainty in the experimental

value of the RME, our predictions for the $A = 8$ systems significantly underestimate the data.

3.2. One- and Two-Body Transition Densities

To investigate the behavior of individual contributions to the RMEs, one- and two-body transition densities are calculated. One can define the one-body transition density as a function of the distance of nucleon i from the center of mass:

$$\mathrm{RME(1b)} = \int \mathrm{d}r_i 4\pi r_i^2 \rho^{1\mathrm{b}}(r_i)\,. \tag{12}$$

In **Figures 2**, **3**, the one-body transition density is found to be consistent between the two cases. Indeed, this is what we expect based on **Table 2**. The LO contribution is consistent between the two interactions and is not sensitive to how the LECs of the three-body interaction are fit. Looking at the sub-leading order contributions, we find that, with the exception of the N2LO-RC term in the $A = 10$ transitions, the only term that is model dependent is the N3LO contact term. To better understand this difference, in a fashion similar to what is done in Equation (12) for the one-body case, a two-body transition density as a function of inter-particle spacing r_{ij} can be defined:

$$\mathrm{RME(2b)} = \int \mathrm{d}r_{ij} 4\pi r_{ij}^2 \rho^{2\mathrm{b}}(r_{ij}) \tag{13}$$

Two-body transition densities are plotted in **Figure 4**. For the same transition, the N2LO-Δ and the N3LO-OPE contributions in models NV2+3-Ia and NV2+3-Ia* are nearly identical. This is to be expected, as the two models use the same cutoffs to regularize the interactions and have the two-body interaction fit to the same data. Where the models differ is in the N3LO-CT contribution. Model NV2+3 Ia* has a larger contribution from this term compared to its counterpart. This model-dependence of the contact contribution to the RME makes sense in light of the difference between models NV2+3-Ia and NV2+3-Ia*. The LECs c_D and c_E entering the three-body contact current were fit with two different procedures. In model NV2+3 Ia, the contribution was constrained using only strong interaction data while in model NV2+3 Ia*, both strong and electroweak data were used to constrain it. This results in the two models having different values for these LECs and thus different strengths in the contact term. While there is evidently a model-dependence, it is worth noting that this is a small contribution to the overall RME for the transitions.

Although there is a model dependence in the N3LO-CT term, it is interesting to ask if the behavior of the current is still similar between the two models. For this purpose, in **Figure 5**, transition densities for the N3LO-CT current selected nuclei are scaled to peak at 1.0 to see if there is a universal behavior in the interactions. The scaling factors to generate **Figure 5** are given in **Table 3**. While there was a difference seen in the size of the contribution of the N3LO-CT term when comparing the unscaled transition densities, it is seen here that the scaled curve overlaps for not only both models within the same transition but also for all transitions under study. The change in the LECs c_D and c_E results in a re-scaling of the N3LO-CT term. For the $A \leq 7$ transitions, the enhancement of the N3LO-CT term relative in NV2+3-Ia* relative to the value given by its counterpart NV2+3-Ia is a factor of ≈ 6.4 for the same transition. For the $A \geq 8$ cases, this enhancement is on average a factor of ≈ 2.2.

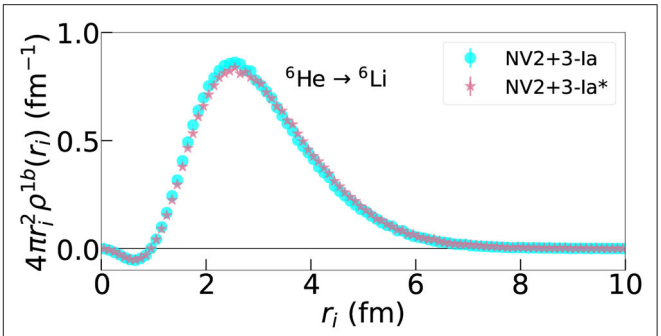


FIGURE 2 | One-body density as defined in Equation (12) for the ^{6}He → ^{6}Li GT RME obtained using the NV2+3-Ia/Ia* models.

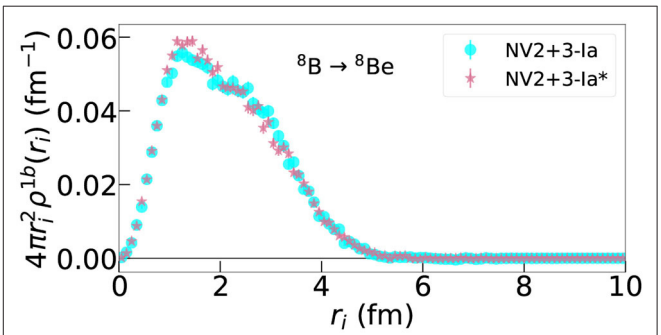


FIGURE 3 | One-body density as defined in Equation (12) for the ^{8}B → ^{8}Be GT RME obtained using the NV2+3-Ia/Ia* models.

Another important feature of the two-body transition densities is the difference in the long-range behavior of the N2LO-Δ and N3LO-OPE terms. In [27], Equations (2.9) and (2.10) give the operator structure of these two currents. In the limit of vanishing momentum transfer, these currents have the same operator structure up to a momentum dependent term that has been verified numerically to provide small contributions. The structures of these operators result in cancelations that are sensitive to the LECs entering into each of these currents, impacting the behavior of the two-body transition density for $r_{ij} \gtrsim 2$ fm. In particular, the N2LO-Δ density is sensitive to the transition under study. In the case of the $A = 10$ transition, the N2LO-Δ density becomes negative near ≈ 2 fm. When integrating over the whole two-body contribution for the NV2+3-Ia* model, this results in a non-trivial cancelation leading to the quenching of the RME from the inclusion of sub-leading contributions.

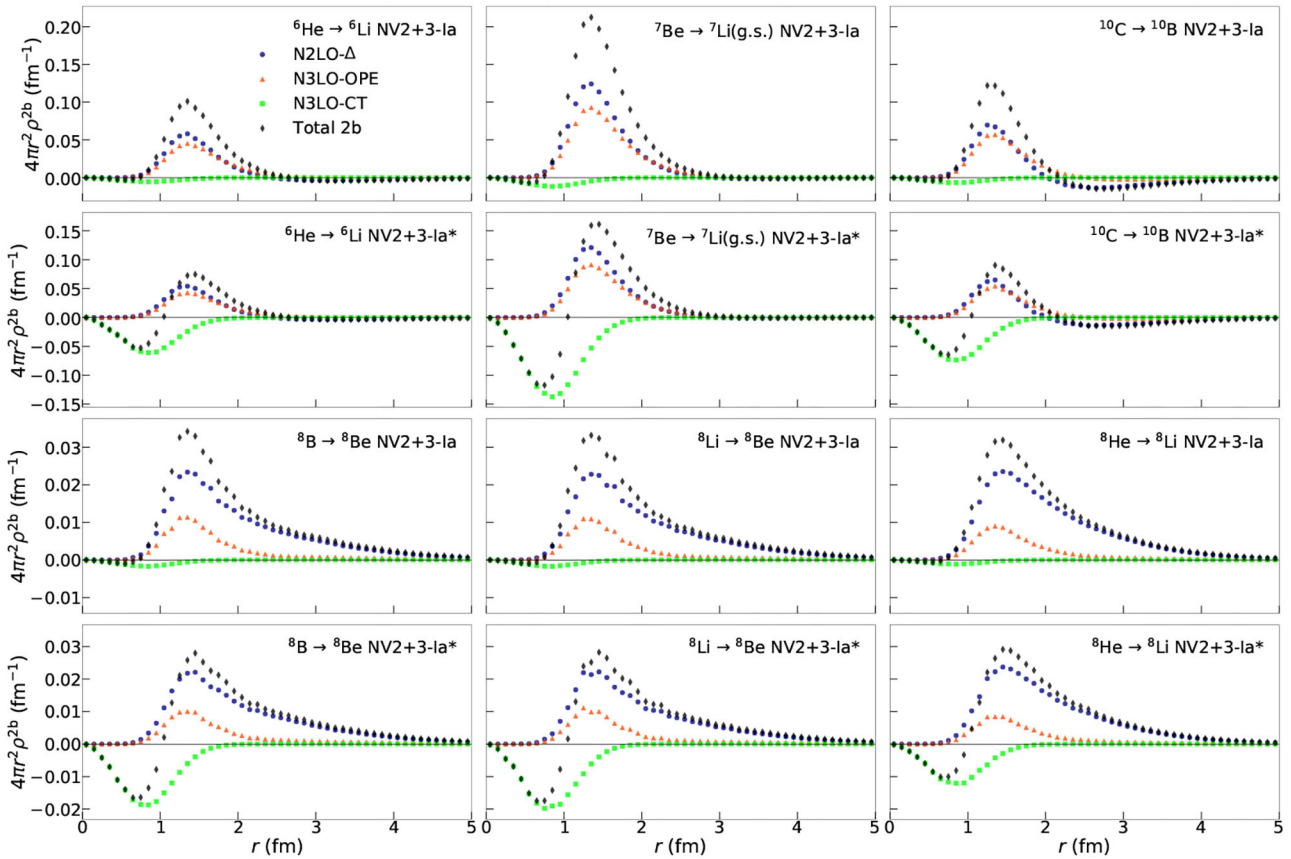


FIGURE 4 | Two-body density as defined in Equation (13) for GT RMEs obtained using the NV2+3-Ia/Ia* models for select nuclei.

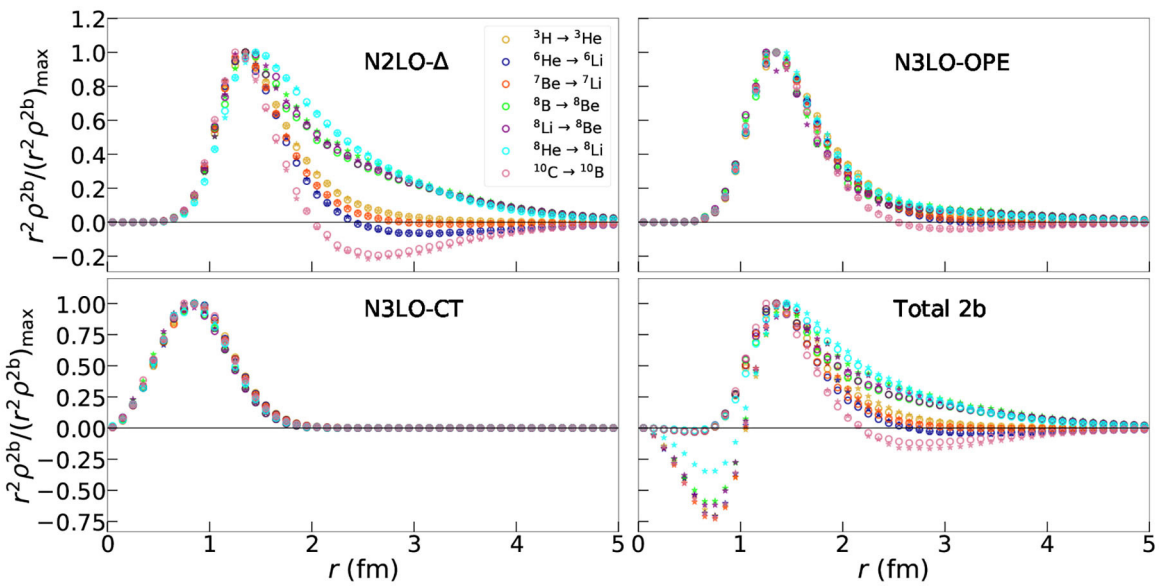


FIGURE 5 | Scaled two-body density as defined in Equation (13) for GT RMEs obtained using the NV2+3-Ia/Ia* models for select nuclei. The distributions are scaled to have a peak value of 1.0 using the factors in **Table 3**.

3.3. GFMC Extrapolation

In addition to VMC calculations of the RME, we also perform a GFMC extrapolation of the RME. The GFMC wave functions generated with the NV2+3-Ia model in this work produce energies that are in statistical agreement with the results of [47]. For all cases presented below, with the exception of two transitions, GFMC propagation are performed between imaginary time steps $\tau = 0.2$ and 0.82 MeV^{-1}. Typically, an imaginary time evolution of the VMC estimate produces an RME that is reduced by a few percent and stabilizes at $\tau \approx 0.2$ MeV^{-1}.

TABLE 3 | The scaling factors $(r_{ij}^2 \rho^{2b})_{max}$ used to normalize the N2LO-Δ, N3LO-OPE, N3LO-CT, and total two-body transition densities in **Figure 5**.

Transition	Model	N2LO-Δ	N3LO-OPE	N3LO-CT	Total 2b
^{3}H → ^{3}He	Ia	0.111	0.082	−0.010	0.189
	Ia*	0.107	0.081	−0.120	0.147
^{6}He → ^{6}Li	Ia	0.058	0.045	−0.005	0.101
	Ia*	0.054	0.043	−0.061	0.075
^{7}Be → ^{7}Li	Ia	0.124	0.093	−0.012	0.212
	Ia*	0.121	0.091	−0.137	0.161
^{8}B → ^{8}Be	Ia	0.023	0.011	−0.002	0.034
	Ia*	0.022	0.010	−0.019	0.028
^{8}Li → ^{8}Be	Ia	0.023	0.011	−0.002	0.033
	Ia*	0.022	0.011	−0.020	0.028
^{8}He → ^{8}Li	Ia	0.023	0.009	−0.001	0.032
	Ia*	0.037	0.008	−0.012	0.029
^{10}C → ^{10}B	Ia	0.070	0.057	−0.006	0.122
	Ia*	0.061	0.050	−0.073	0.085

See text for details.

In the calculations of transitions involving the $(J^\pi; T) = (2^+; 0)$ state of ^{8}Be and the ground state of ^{8}B, the extrapolation must be treated differently. In this two states, as τ increases, the binding energy, magnitude of the quadrupole moment, and point-proton radius all increase monotonically. This is interpreted as the dissolution of ^{8}Be into two alpha particles and ^{8}B into p+^{7}Be. Datar et al. [71], Pastore et al. [72], and Wiringa et al. [89] have previously addressed this issue for ^{8}Be. Similar to those references, we perform the extrapolation by noting that the energy drops rapidly in τ, stabilizing at $\tau \approx 0.1$ MeV^{-1}. We assume that, at this point, spurious contamination in the wave function has been removed by the GFMC procedure and average in a small interval around $\tau = 0.1$ MeV^{-1}, taken to be τ from 0.06 to 0.14 MeV^{-1}. This introduces and additional ≈ 5% systematic uncertainty to these calculations in addition to the statistical uncertainties of QMC.

In all transitions, except for the NV2+3-Ia* model for $A = 10$, the GFMC extrapolation reduces the VMC RME by a $\lesssim$ 4%. **Table 4** summarizes results of the LO, total sub-leading order (Total-LO), and total RMEs for the transitions under study. In the $A < 10$ transitions, the LO contribution is consistent between the two different models under study. In the $A = 10$ case, this model dependence can be understood by the existence of nearby $J^\pi = 1^+$ excited states in ^{10}B. The lower state is predominantly ^{3}S$_1$[442] state and the upper one a ^{3}D$_1$[442] state. These two states are split by only 1 MeV. The transition from the predominantly ^{1}S$_0$[442] ^{10}C(0$^+$) state is large in the $S \rightarrow S$ components but about five times smaller in the $S \rightarrow D$ components. This causes the GT matrix element to be particularly sensitive to the exact mixing of the ^{3}S$_1$ and ^{3}D$_1$ components in the two ^{10}B(1$^+$) states produced by a given Hamiltonian, as was observed for the calculation of GT matrix elements using the AV18+IL7 interaction [73]. In either case, the NV2+3 interactions overpredict the data.

TABLE 4 | Gamow-Teller RMEs in A = 6, 7, 8, and 10 nuclei obtained with chiral axial currents [27] and GFMC (VMC) wave functions corresponding to the NV2+3-Ia/Ia* Hamiltonian models [24, 25, 27, 47].

Transition	Model	LO	(Total-LO)	Total	Expt.
^{6}He(0$^+$;1)→ ^{6}Li(1$^+$;0)	Ia	2.125 (2.200)	0.071 (0.056)	2.195 (2.256)	2.1609 (40)
[42] → [42]	Ia*	2.107 (2.192)	0.011 (0.005)	2.118 (2.197)	
^{7}Be($\frac{3}{2}^-$;$\frac{1}{2}$)→ ^{7}Li($\frac{3}{2}^-$;$\frac{1}{2}$)	Ia	2.273 (2.317)	0.164 (0.165)	2.440 (2.482)	2.3556 (47)
[43] → [43]	Ia*	2.286 (2.327)	0.052 (0.053)	2.338 (2.380)	
^{7}Be($\frac{3}{2}^-$;$\frac{1}{2}$)→ ^{7}Li($\frac{1}{2}^-$;$\frac{1}{2}$)	Ia	2.065 (2.157)	1.03 (0.121)	2.168 (2.278)	2.1116 (57)
[43] → [43]	Ia*	2.061 (2.158)	0.009 (0.025)	2.070 (2.183)	
^{8}Li(2$^+$;1)→ ^{8}Be(2$^+$;0)	Ia	0.074 (0.147)	0.029 (0.041)	0.103 (0.188)	0.284 [84]
[431] → [44]	Ia*	0.096 (0.148)	0.025 (0.026)	0.120 (0.174)	0.190 [85]
^{8}B(2$^+$;1)→ ^{8}Be(2$^+$;0)	Ia	0.091 (0.146)	0.035 (0.042)	0.125 (0.188)	0.269 (20)
[431] → [44]	Ia*	0.102 (0.148)	0.024 (0.026)	0.126 (0.174)	
^{8}He(0$^+$;2)→ ^{8}Li(1$^+$;1)	Ia	0.262 (0.386)	0.040 (0.038)	0.302 (0.424)	0.512 (6)
[422] → [431]	Ia*	0.297 (0.362)	0.025 (0.029)	0.322 (0.391)	
^{10}C(0$^+$;1)→ ^{10}B(1$^+$;0)	Ia	1.928 (1.940)	0.050 (0.041)	1.978 (1.981)	1.8331 (34)
[442] → [442]	Ia*	2.086 (2.015)	−0.031 (−0.037)	2.055 (1.978)	

Results corresponding to the one-body current at LO (column labeled "LO"), and to the sum of all the corrections beyond LO (column labeled "Total-LO") are given, along with the cumulative contributions (column labeled "Total") to be compared with the experimental data [81–85] reported in the last row. Statistical errors associated with the Monte Carlo integrations are not shown, but are below 1%. Transitions for the A = 8 systems are affected by an additional systematic error of ∼ %, see text for explanation.

4. CONCLUSIONS

With this work we set the foundations for the development of an accurate and unified understanding of neutrino–nucleus interactions. We are in the process of exploring and validating the QMC approach's description of electroweak processes in a wide range of energy and momentum transfer; in this work, we therefore focused on calculating matrix elements entering beta decay and inverse beta decay in light nuclei. These processes occur at zero momentum transfer and involve energy transferred of the order of a few MeVs.

In our approach, we fully retained two- and three-nucleon correlations induced by the Norfolk potentials, and we described the interaction with the external electroweak probes by means of the associated one- and two-body axial currents at tree-level. This study was focused on the NV2+3-Ia and NV2+3-Ia* models, and was aimed at carefully studying the contributions from two-body axial currents in the two different implementations of the three-nucleon forces. In the unstarred model the LECs c_D and c_E entering the three-nucleon force were fitted to the trinucleon binding energies and the *nd* scattering length, while the starred model is constrained by the experimental GT value of the triton decay and the trinucleon binding energies. The axial two-body contact current at N3LO, which involves c_D, was taken consistently with the three-nucleon force adopted to generate the nuclear wave functions.

In analogy with previous QMC studies of beta decay in light nuclei [36, 73], we find that corrections from two-body axial currents are at the $\sim$ 3% level in $A = 6, 7$ and 10. The $A = 8$ systems are instead severely underpredicted by the theory, even after the inclusion of large ($\sim 30-40\%$) contributions from two-body axial currents. Studies on the electromagnetic transitions in $A = 8$ nuclei were also found to be problematic [70, 72], which indicates the need of further developments of the $A = 8$ wave functions.

In this work, we especially focused on the contributions from two-body currents, which, despite the fact that they are in these cases small, can provide us with valuable insights on the composition of these corrections. To this end, we reported studies on two-body transition densities which allow us to understand the relevance of the two-body currents as a function of the interparticle distance. As expected, we find that, for a given interaction model, the transition densities exhibit a universal short-range behavior across the considered nuclei, while they differ in the long-range tails. The starred and unstarred results differ in the contact contribution at N3LO. This is rather visible in the panels of **Figure 4** where the starred model leads to a total transition density (black symbols) which presents one or more nodes. The presence of nodes implies non-trivial cancelations when using the starred models.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This research was supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, under the U.S. Department of Energy funds through the FRIB Theory Alliance award DE-SC0013617 (MP and SP). LA and SP have been supported by the U.S. Department of Energy under contract DE-SC0021027.

ACKNOWLEDGMENTS

We thank our collaborators J. Carlson, S. Gandolfi, R. Schiavilla, and R. B. Wiringa for their contributions to the studies presented in this work. The many-body calculations were performed on the parallel computers of the Laboratory Computing Resource Center, Argonne National Laboratory, the computers of the Argonne Leadership Computing Facility (ALCF) via the ALCC-2019 grant Low energy neutrino-nucleus interactions.

REFERENCES

1. Gando A, Gando Y, Hachiya T, Hayashi A, Hayashida S, Ikeda H, et al. Publisher's note: search for majorana neutrinos near the inverted mass hierarchy region with KamLAND-Zen [Phys. Rev. Lett. 117, 082503 (2016)]. *Phys Rev Lett*. (2016) **117**:109903. doi: 10.1103/PhysRevLett.117.082503
2. Albert JB, Auty DJ, Barbeau P, Beauchamp E. Search for Majorana neutrinos with the first two years of EXO-200 data. *Nature*. (2014) **510**:229–34. doi: 10.1038/nature13432
3. Adams C, Alvarez V, Arazi L, Arnquist IJ, Azevedo CDR, Bailey K, et al. Sensitivity of a tonne-scale NEXT detector for neutrinoless double beta decay searches. *arXiv [Preprint] arXiv*:2005.06467. (2020)
4. Zsigmond AJ. LEGEND: The future of neutrinoless double-beta decay search with germanium detectors. *J Phys Conf Ser*. (2020) **1468**:012111. doi: 10.1088/1742-6596/1468/1/012143
5. Cattadori CM. Gerda: results and perspectives. *Nucl Part Phys Proc*. (2015) **265–6**:38–41. doi: 10.1016/j.nuclphysbps.2015.06.010
6. Alfonso K, Artusa DR, Avignone FT, Azzolini O, Balata M, Banks TI, et al. Search for neutrinoless double-beta decay of ^{1}30Te with CUORE-0. *Phys Rev Lett*. (2015) **115**:102502. doi: 10.1103/PhysRevLett.115.102502
7. Caden E. Status of the SNO+ experiment. *J Phys Conf Ser*. (2020) **1342**:012022. doi: 10.1088/1742-6596/1342/1/012022
8. Blot S. Investigating $\beta\beta$ decay with the NEMO-3 and SuperNEMO experiments. *J Phys Conf Ser*. (2016) **718**:062006. doi: 10.1088/1742-6596/718/6/062006
9. Giuliani A. A neutrinoless double-beta-decay search based on $ZnMoO_4$ and Li_2MoO_4 scintillating bolometers. *J Phys Conf Ser*. (2017) **888**:012239. doi: 10.1088/1742-6596/888/1/012239
10. Tetsuno K, Ajimura S, Akutagawa K, Batpurev T, Chan WM, Fushimi K, et al. Status of 48Ca double beta decay search and its future prospect in CANDLES. *J Phys Conf Ser*. (2020) **1468**:012132. doi: 10.1088/1742-6596/1468/1/012132

11. Park H. The AMoRE: Search for neutrinoless double beta decay in 100Mo. *Nucl Part Phys Proc*. (2016) **273–5**:2630–2. doi: 10.1016/j.nuclphysbps.2015.10.012
12. Ebert J, Fritts M, Gehre D, Gößling C, Göpfert T, Hagner C, et al. The COBRA demonstrator at the LNGS underground laboratory. *Nucl Instrum Meth A*. (2016) **807**:114–20. doi: 10.1016/j.nima.2015.10.079
13. Dokania N, Singh V, Ghosh C, Mathimalar S, Garai A, Pal S, et al. Radiation background studies for $0\nu\beta\beta$ decay in ^{1}24Sn. In: *Topical Research Meeting on Prospects in Neutrino Physics*. London, UK (2015).
14. Fukuda Y. ZICOS - New project for neutrinoless double beta decay experiment using zirconium complex in liquid scintillator. *J Phys*. (2016) **718**:062019. doi: 10.1088%2F1742-6596%2F718%2F6%2F062019
15. Engel J, Menndez J. Status and future of nuclear matrix elements for neutrinoless double-beta decay: a review. *Rept Prog Phys*. (2017) **80**:046301. doi: 10.1088/1361-6633/aa5bc5
16. [Dataset] The MicroBooNE Experiment. Available online at: http://www-microboone.fnal.gov
17. [Dataset] The T2K Experiment. Available online at: http://t2k-experiment.org
18. [Dataset] The Deep Underground Neutrino Experiment. Available online at: http://www.dunescience.org
19. [Dataset] The NOvA Experiment. Available online at: http://www-nova.fnal.gov
20. [Dataset] Hyper-Kamiokande. Available online at: http://www.hyperk.org
21. Carlson J, Gandolfi S, Pederiva F, Pieper SC, Schiavilla R, Schmidt KE, et al. Quantum Monte Carlo methods for nuclear physics. *Rev Mod Phys*. (2015) **87**:1067. doi: 10.1103/RevModPhys.87.1067
22. Lynn JE, Tews I, Gandolfi S, Lovato A. Quantum Monte Carlo methods in nuclear physics: recent advances. *Ann Rev Nucl Part Sci*. (2019) **69**:279–305. doi: 10.1146/annurev-nucl-101918-023600
23. Gandolfi S, Lonardoni D, Lovato A, Piarulli M. Atomic nuclei from quantum Monte Carlo calculations with chiral EFT interactions. (2020) *Front Phys*. **8**:117. doi: 10.3389/fphy.2020.00117
24. Piarulli M, Girlanda L, Schiavilla R, Navarro Pérez R, Amaro JE, Ruiz Arriola E. Minimally nonlocal nucleon-nucleon potentials with chiral two-pion exchange including Δ resonances. *Phys Rev C*. (2015) **91**:024003. doi: 10.1103/PhysRevC.91.024003
25. Piarulli M, Girlanda L, Schiavilla R, Kievsky A, Lovato A, Marcucci LE, et al. Local chiral potentials with Δ-intermediate states and the structure of light nuclei. *Phys Rev C*. (2016) **94**:054007. doi: 10.1103/PhysRevC.94.054007
26. Baroni A, Girlanda L, Kievsky A, Marcucci LE, Schiavilla R, Viviani M. Tritium β-decay in chiral effective field theory. *Phys Rev C*. (2016) **94**:024003. doi: 10.1103/PhysRevC.94.024003
27. Baroni A, Schiavilla R, Marcucci LE, Girlanda L, Kievsky A, Lovato A, et al. Local chiral interactions, the tritium Gamow-Teller matrix element, and the three-nucleon contact term. *Phys Rev C*. (2018) **98**:044003. doi: 10.1103/PhysRevC.98.044003
28. Piarulli M, Tews I. Local nucleon-nucleon and three-nucleon interactions within chiral effective field theory. *Front Phys*. (2020) **7**:245. doi: 10.3389/fphy.2019.00245
29. Baroni A, Girlanda L, Pastore S, Schiavilla R, Viviani M. Nuclear axial currents in chiral effective field theory. *Phys Rev C*. (2016) **93**:015501. doi: 10.1103/PhysRevC.93.049902
30. Coraggio L, Gargano A, Itaco N, Mancino R, Nowacki F. The calculation of the neutrinoless double-beta decay matrix element within the realistic shell model. *Phys Rev C*. (2020) **101**:044315. doi: 10.1103/PhysRevC.101.044315
31. Coraggio L, Itaco N, Mancino R. Short-range correlations for neutrinoless double-beta decay and low-momentum NN potentials. In: *27th International Nuclear Physics Conference*. Glasgow, UK (2019).
32. Wang XB, Hayes AC, Carlson J, Dong GX, Mereghetti E, Pastore S, et al. Comparison between variational Monte Carlo and shell model calculations of neutrinoless double beta decay matrix elements in light nuclei. *Phys Lett B*. (2019) **798**:134974. doi: 10.1016/j.physletb.2019.134974
33. Pastore S, Carlson J, Gandolfi S, Schiavilla R, Wiringa RB. Quasielastic lepton scattering and back-to-back nucleons in the short-time approximation. *Phys Rev C*. (2019) **101**:044612. doi: 10.1103/PhysRevC.101.044612
34. Lovato A, Carlson J, Gandolfi S, Rocco N, Schiavilla R. *Ab initio* study of (ν_ℓ, ℓ^-) and $(\overline{\nu}_\ell, \ell^+)$ inclusive scattering in ^{1}2C: confronting the MiniBooNE and T2K CCQE data. *Phys Rev X*. (2020) **3**:031068. doi: 10.1103/PhysRevX.10.031068
35. Alvarez-Ruso L, Athar MS, Barbaro MB, Cherdack D, Christy ME, Coloma P, et al. NuSTEC1 1Neutrino scattering theory experiment collaboration http://nustec.fnal.gov. White Paper: Status and challenges of neutrino–nucleus scattering. *Prog Part Nucl Phys*. (2018) **100**:1–68. doi: 10.1016/j.ppnp.2018.01.006
36. King GB, Andreoli L, Pastore S, Piarulli M, Schiavilla R, Wiringa RB, et al. Chiral effective field theory calculations of weak transitions in light nuclei. *Phys Rev C*. (2020) **102**:025501. doi: 10.1103/PhysRevC.102.025501
37. Wiringa RB. Variational calculations of few-body nuclei. *Phys Rev C*. (1991) **43**:1585–98. doi: 10.1103/PhysRevC.43.1585
38. Pudliner BS, Pandharipande VR, Carlson J, Pieper SC, Wiringa RB. Quantum Monte Carlo calculations of nuclei with A <= 7. *Phys Rev C*. (1997) **56**:1720–50. doi: 10.1103/PhysRevC.56.1720
39. Nollett KM, Wiringa RB, Schiavilla R. A Six body calculation of the alpha deuteron radiative capture cross-section. *Phys Rev C*. (2001) **63**:024003. doi: 10.1103/PhysRevC.63.024003
40. Nollett KM. Radiative alpha capture cross-sections from realistic nucleon-nucleon interactions and variational Monte Carlo wave functions. *Phys Rev C*. (2001) **63**:054002. doi: 10.1103/PhysRevC.63.054002
41. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *J Chem Phys*. (1953) **21**:1087–92. doi: 10.1063/1.1699114
42. Pervin M, Pieper SC, Wiringa RB. Quantum Monte Carlo calculations of electroweak transition matrix elements in A = 6,7 nuclei. *Phys Rev C*. (2007) **76**:064319. doi: 10.1103/PhysRevC.76.064319
43. Krebs H, Epelbaum E, Meissner UG. Nuclear forces with Delta-excitations up to next-to-next-to-leading order. I. Peripheral nucleon-nucleon waves. *Eur Phys J A*. (2007) **32**:127–37. doi: 10.1140/epja/i2007-10372-y
44. Navarro Pérez R, Amaro JE, Ruiz Arriola E. Coarse-grained potential analysis of neutron-proton and proton-proton scattering below the pion production threshold. *Phys Rev C*. (2013) **88**:064002. doi: 10.1103/PhysRevC.88.064002
45. Navarro Pérez R, Amaro JE, Ruiz Arriola E. Coarse grained NN potential with chiral two pion exchange. *Phys Rev C*. (2014) **89**:024004. doi: 10.1103/PhysRevC.89.024004
46. Navarro Perez R, Amaro JE, Ruiz Arriola E. Statistical error analysis for phenomenological nucleon-nucleon potentials. *Phys Rev C*. (2014) **89**:064006. doi: 10.1103/PhysRevC.89.064006
47. Piarulli M, Baroni A, Girlanda L, Kievsky A, Lovato A, Lusk E, et al. Light-nuclei spectra from chiral dynamics. *Phys Rev Lett*. (2018) **120**:052503. doi: 10.1103/PhysRevLett.120.052503
48. van Kolck U. Few nucleon forces from chiral Lagrangians. *Phys Rev C*. (1994) **49**:2932–41. doi: 10.1103/PhysRevC.49.2932
49. Epelbaum E, Nogga A, Gloeckle W, Kamada H, Meissner UG, Witala H. Three nucleon forces from chiral effective field theory. *Phys Rev C*. (2002) **66**:064001. doi: 10.1103/PhysRevC.66.064001
50. Lynn JE, Tews I, Carlson J, Gandolfi S, Gezerlis A, Schmidt KE, et al. Chiral three-nucleon interactions in light nuclei, neutron-α scattering, and neutron matter. *Phys Rev Lett*. (2016) **116**:062501. doi: 10.1103/PhysRevLett.116.062501
51. Tews I, Gandolfi S, Gezerlis A, Schwenk A. Quantum Monte Carlo calculations of neutron matter with chiral three-body forces. *Phys Rev C*. (2016) **93**:024305. doi: 10.1103/PhysRevC.93.024305
52. Lynn JE, Tews I, Carlson J, Gandolfi S, Gezerlis A, Schmidt KE, et al. Quantum Monte Carlo calculations of light nuclei with local chiral two- and three-nucleon interactions. *Phys Rev C*. (2017) **96**:054007. doi: 10.1103/PhysRevC.96.054007
53. Gazit D, Quaglioni S, Navratil P. Three-nucleon low-energy constants from the consistency of interactions and currents in chiral effective field theory. *Phys Rev Lett*. (2009) **103**:102502. doi: 10.1103/PhysRevLett.103.102502
54. Marcucci LE, Kievsky A, Rosati S, Schiavilla R, Viviani M. Chiral effective field theory predictions for muon capture on deuteron and ^{3}*He*. *Phys Rev Lett*. (2012) **108**:052502. doi: 10.1103/PhysRevLett.108.052502
55. Gardestig A, Phillips DR. How low-energy weak reactions can constrain three-nucleon forces and the neutron-neutron scattering length. *Phys Rev Lett*. (2006) **96**:232301. doi: 10.1103/PhysRevLett.96.232301

56. Lovato A, Benhar O, Fantoni S, Schmidt KE. Comparative study of three-nucleon potentials in nuclear matter. *Phys Rev C*. (2012) **85**:024003. doi: 10.1103/PhysRevC.85.024003
57. Cirigliano V, Dekens W, De Vries J, Graesser ML, Mereghetti E, Pastore S, et al. A renormalized approach to neutrinoless double-beta decay. *Phys Rev C*. (2019) **100**:055504. doi: 10.1103/PhysRevC.100.055504
58. Piarulli M, Bombaci I, Logoteta D, Lovato A, Wiringa RB. Benchmark calculations of pure neutron matter with realistic nucleon-nucleon interactions. *Phys Rev C*. (2019) **101**:045801. doi: 10.1103/PhysRevC.101.045801
59. Demorest P, Pennucci T, Ransom S, Roberts M, Hessels J. Shapiro delay measurement of a two solar mass neutron star. *Nature*. (2010) **467**:1081–3. doi: 10.1038/nature09466
60. Antoniadis J, Freire PCC, Wex N, Tauris TM, Lynch RS, van Kerkwijk MH, et al. A massive pulsar in a compact relativistic binary. *Science*. (2013) **340**:6131. doi: 10.1126/science.1233232
61. Bacca S, Pastore S. Electromagnetic reactions on light nuclei. *J Phys G*. (2014) **41**:123002. doi: 10.1088/0954-3899/41/12/123002
62. Piarulli M, Girlanda L, Marcucci LE, Pastore S, Schiavilla R, Viviani M. Electromagnetic structure of A = 2 and 3 nuclei in chiral effective field theory. *Phys Rev C*. (2013) **87**:014006. doi: 10.1103/PhysRevC.87.014006
63. Schiavilla R, Baroni A, Pastore S, Piarulli M, Girlanda L, Kievsky A, et al. Local chiral interactions and magnetic structure of few-nucleon systems. *Phys Rev C*. (2019) **99**:034005. doi: 10.1103/PhysRevC.99.034005
64. Nevo Dinur N, Hernandez OJ, Bacca S, Barnea N, Ji C, Pastore S, et al. Zemach moments and radii of ^{2}, 3H and ^{3}, 4He. *Phys Rev C*. (2019) **99**:034004. doi: 10.1103/PhysRevC.99.034004
65. Marcucci LE, Gross F, Pena MT, Piarulli M, Schiavilla R, Sick I, et al. Electromagnetic structure of few-nucleon ground states. *J Phys G*. (2016) **43**:023002. doi: 10.1088/0954-3899/43/2/023002
66. Pastore S, Schiavilla R, Goity JL. Electromagnetic two-body currents of one- and two-pion range. *Phys Rev C*. (2008) **78**:064002. doi: 10.1103/PhysRevC.78.064002
67. Pastore S, Girlanda L, Schiavilla R, Viviani M, Wiringa RB. Electromagnetic currents and magnetic moments in (chi)EFT. *Phys Rev C*. (2009) **80**:034004. doi: 10.1103/PhysRevC.80.034004
68. Girlanda L, Kievsky A, Marcucci LE, Pastore S, Schiavilla R, Viviani M. Thermal neutron captures on d and ^{3}He. *Phys Rev Lett*. (2010) **105**:232502. doi: 10.1103/PhysRevLett.105.232502
69. Pastore S, Girlanda L, Schiavilla R, Viviani M. The two-nucleon electromagnetic charge operator in chiral effective field theory (χEFT) up to one loop. *Phys Rev C*. (2011) **84**:024001. doi: 10.1103/PhysRevC.84.024001
70. Pastore S, Pieper SC, Schiavilla R, Wiringa RB. Quantum Monte Carlo calculations of electromagnetic moments and transitions in A $\leq$ 9 nuclei with meson-exchange currents derived from chiral effective field theory. *Phys Rev C*. (2013) **87**:035503. doi: 10.1103/PhysRevC.87.035503
71. Datar VM, Chakrabarty DR, Suresh Kumar, Nanal V, Pastore S, Wiringa RB, et al. Electromagnetic transition from the 4+ to 2+ resonance in Be8 measured via the radiative capture in He4+He4. *Phys Rev Lett*. (2013) **111**:062502. doi: 10.1103/PhysRevLett.111.062502
72. Pastore S, Wiringa RB, Pieper SC, Schiavilla R. Quantum Monte Carlo calculations of electromagnetic transitions in ^{8}Be with meson-exchange currents derived from chiral effective field theory. *Phys Rev C*. (2014) **90**:024321. doi: 10.1103/PhysRevC.90.024321
73. Pastore S, Baroni A, Carlson J, Gandolfi S, Pieper SC, Schiavilla R, et al. Quantum Monte Carlo calculations of weak transitions in $A = 6 - 10$ nuclei. *Phys Rev C*. (2018) **97**:022501. doi: 10.1103/PhysRevC.97.022501
74. Pastore S, Carlson J, Cirigliano V, Dekens W, Mereghetti E, Wiringa RB. Neutrinoless double-β decay matrix elements in light nuclei. *Phys Rev C*. (2018) **97**:014606. doi: 10.1103/PhysRevC.97.014606
75. Cirigliano V, Dekens W, De Vries J, Graesser ML, Mereghetti E, Pastore S, et al. New leading contribution to neutrinoless double-β decay. *Phys Rev Lett*. (2018) **120**:202001. doi: 10.1103/PhysRevLett.120.202001
76. Krebs H, Epelbaum E, Meißner UG. Box diagram contribution to the axial two-nucleon current. *Phys Rev C*. (2020) **101**:055502. doi: 10.1103/PhysRevC.101.055502
77. Park TS, Min DP, Rho M. Chiral dynamics and heavy fermion formalism in nuclei. 1. Exchange axial currents. *Phys Rept*. (1993) **233**:341–95. doi: 10.1016/0370-1573(93)90099-Y
78. Kolling S, Epelbaum E, Krebs H, Meissner UG. Two-pion exchange electromagnetic current in chiral effective field theory using the method of unitary transformation. *Phys Rev C*. (2009) **80**:045502. doi: 10.1103/PhysRevC.80.045502
79. Kolling S, Epelbaum E, Krebs H, Meissner UG. Two-nucleon electromagnetic current in chiral effective field theory: one-pion exchange and short-range contributions. *Phys Rev C*. (2011) **84**:054008. doi: 10.1103/PhysRevC.84.054008
80. Tanabashi M, Hagiwara K, Hikasa K, Nakamura K, Sumino Y, Takahashi F, et al. Review of particle physics. *Phys Rev D*. (2018) **98**:030001. doi: 10.1103/PhysRevD.98.030001
81. Knecht A, Hong R, Zumwalt DW, Delbridge BG, Garcia A. Precision measurement of the He-6 half-life and the weak axial current in nuclei. *Phys Rev C*. (2012) **86**:035506. doi: 10.1103/PhysRevC.86.035506
82. Suzuki T, Fujimoto R, Otsuka T. Gamow-Teller transitions and magnetic properties of nuclei and shell evolution. *Phys Rev C*. (2003) **67**:044302. doi: 10.1103/PhysRevC.67.044302
83. Chou WT, Warburton EK, Brown BA. Gamow-teller beta-decay rates for A <= 18 nuclei. *Phys Rev C*. (1993) **47**:163–77. doi: 10.1103/PhysRevC.47.163
84. Warburton EK. R-matrix analysis of the $\beta^{\mp}$-delayed alpha spectra from the decay of ^{8}Li and ^{8}B. *Phys Rev C*. (1986) **33**:303–13. doi: 10.1103/PhysRevC.33.303
85. Pritychenko B, BK E, Kellett MA, Singh B, Totans J. The Nuclear Science References (NSR) database and Web Retrieval System. *Nucl Instrum Methods Phys Res Sect A*. (2011) **640**:213–8. doi: 10.1016/j.nima.2011.03.018
86. Wiringa RB. Pair counting, pion-exchange forces, and the structure of light nuclei. *Phys Rev C*. (2006) **73**:034317. doi: 10.1103/PhysRevC.73.034317
87. Gysbers P, Hagen G, Holt JD, Jansen GR, Morris TD, Navratil P, et al. Discrepancy between experimental and theoretical β-decay rates resolved from first principles. *Nat Phys*. (2019) **15**:428–31. doi: 10.1038/s41567-019-0450-7
88. Hardy JC, Towner IS. Superallowed $0^+ \rightarrow 0^+$ nuclear decays: 2014 critical survey, with precise results for $V_u d$ and CKM unitarity. *Phys Rev C*. (2015) **91**:025501. doi: 10.1103/PhysRevC.91.025501
89. Wiringa RB, Pieper SC, Carlson J, Pandharipande VR. Quantum Monte Carlo calculations of A = 8 nuclei. *Phys Rev C*. (2000) **62**:014001. doi: 10.1103/PhysRevC.62.014001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 26 February 2020
doi: 10.3389/fphy.2020.00036

Reinterpretation of Classic Proton Charge Form Factor Measurements

Miha Mihovilovič [1,2,3], Douglas W. Higinbotham [4], Melisa Bevc [1] and Simon Širca [1,2]*

[1] Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia, [2] Department of Low and Medium Energy Physics, Jožef Stefan Institute, Ljubljana, Slovenia, [3] Institut für Kernphysik, Johannes-Gutenberg-Universität, Mainz, Germany, [4] Thomas Jefferson National Accelerator Facility, Newport News, VA, United States

In 1963, a proton radius of 0.805(11) fm was extracted from electron scattering data and this classic value has been used in the standard dipole parameterization of the form factor. In trying to reproduce this classic result, we discovered that there was a sign error in the original analysis and that the authors should have found a value of 0.851(19) fm. We additionally made use of modern computing power to find a robust function for extracting the radius using this 1963 data's spacing and uncertainty. This optimal function, the Padé (0, 1) approximant, also gives a result which is consistent with the modern high precision proton radius extractions.

Keywords: proton, charge radius, form factors, statistical methods, electron scattering

OPEN ACCESS

Edited by:
Laura Elisa Marcucci, University of Pisa, Italy

Reviewed by:
Mitko Gaidarov, Institute for Nuclear Research and Nuclear Energy (BAS), Bulgaria
Roelof Bijker, National Autonomous University of Mexico, Mexico

****Correspondence:***
Miha Mihovilovič
miha.mihovilovic@ijs.si

Specialty section:
This article was submitted to Nuclear Physics, a section of the journal Frontiers in Physics

Received: *03 December 2019*
Accepted: *05 February 2020*
Published: *26 February 2020*

Citation:
Mihovilovič M, Higinbotham DW, Bevc M and Širca S (2020) Reinterpretation of Classic Proton Charge Form Factor Measurements. Front. Phys. 8:36. doi: 10.3389/fphy.2020.00036

1. INTRODUCTION

The proton charge radius, r_E, is the conventional measure for the size of the proton, a fundamental constituent of matter. This constant is defined as the derivative of the proton charge form factor, G_E^p, at zero four-momentum transfer, $Q^2 = 0$:

$$r_E^2 \equiv -6\hbar^2 \left.\frac{dG_E^p}{dQ^2}\right|_{Q^2=0}, \tag{1}$$

and can be determined by both hydrogen spectroscopy and elastic lepton scattering [1]. The first determination of the radius was done with elastic electron scattering data by Hand et al. [2], who determined the radius of 0.805(11) fm, the value used in the standard dipole parameterization of the form factor [3, 4]. The original study was followed by several decades of dedicated nuclear scattering and spectroscopic experiments, which led to a recommended value for the proton charge radius of 0.8791(79) fm (CODATA 2010, [5]). This result was called into question when the extremely precise spectroscopic measurements on muonic hydrogen [6, 7] reported a significantly smaller value of 0.84087(39) fm. The observed discrepancy, colloquially known as "the proton radius puzzle" [8] motivated several new experiments [9–12]. These experiments have been accompanied by different reanalyses of the existing data [13–20], focusing on data of Bernauer et al. [21, 22]. In this paper we follow a different path and revisit the first data of Hand et al., and evaluate their result by using modern analysis techniques.

2. THE CLASSICAL APPROACH

In the first determination of the radius, existing data on proton charge form factor from five different measurements were considered [23–27], as noted in **Table 1**.

TABLE 1 | Summary of the experimental data considered in the analysis.

References	Number of data points	Q^2_{min} [fm^{-2}]	Q^2_{max} [fm^{-2}]	Average uncertainty
Litauer et al. [23]	4	2.	8.	0.251
Bumiller et al. [24]	10	0.36	10.	0.051
Drickey et al. [25]	4	0.3	2.2	0.006
Yount et al. [26]	3	0.28	1.3	0.016
Lehmann et al. [27]	6	0.3	2.98	0.012

For each data set, the columns represent the number of measured points, the minimal and maximal value of four-momentum transfer at which $G_E^p(Q^2)$ was measured, and the average experimental uncertainty.

In an attempt to reconstruct the radius of 0.81 fm we followed the original analysis approach and compared the data to the quadratic function in Q^2:

$$G_{\text{quadratic}}(Q^2) = 1 - \frac{r_E^2}{6}Q^2 + aQ^4 . \quad (2)$$

This model depends on two free parameters: the radius, r_E, in front of the linear term, and the parameter a that determines the curvature of the function. Since the data are normalized, the constant term of the model is simply 1. In the first step the two parameters were determined by fitting Equation (2) to the data with $Q^2 \leq 3\,\text{fm}^{-2}$, considering the entire region with the high density of experimental points. The obtained results were $r_E = 0.819(21)\,\text{fm}$ and $a = 0.00787(309)\,\text{fm}^4$. However, the radius obtained in this manner should not be trusted since the true shape of the $G_E^p(Q^2)$ may be more complex than a second order polynomial. At $Q^2 \approx 3\,\text{fm}^{-2}$ the contributions of the Q^6 and Q^8 terms are not negligible and their omission from the fit causes a systematic shift in the determined radius.

To avoid model dependent bias in the radius extraction, the contributions of higher order terms should be kept minimal. The way Hand achieved this with a model, such as Equation (2), is by keeping the parameter a at a value determined in their first step and then only fitting the radius, using data with $Q^2 \leq 1.05\,\text{fm}^{-2}$. Assuming that the determined value for a is a good estimate for the size of the Q^4 term, this preserves the curvature of the model. Additionally, we were able to determine that at $1\,\text{fm}^{-2}$ the Q^4 term contributes less than a percent to the value of G_E^p. Hence, even a 10 % error in the value of a would result in a modification of the form-factor much smaller than the statistical uncertainty of each measurement. Hence, the described two step fitting technique should result in a more reliable estimate of the proton charge radius. We determined it to be $r_E = 0.851(19)$ fm, which is inconsistent with the original result (see **Figure 1**). The obtained value is 5% larger than the original radius while its uncertainty is almost twice as large as the uncertainty of the first result.

To find the source of the discrepancy the last step of the analysis was repeated with different values of a. Since r_E and a are strongly correlated, it is important to evaluate the effect of a on r_E. Additionally, the original paper does not report the value of a. The analysis demonstrated in **Figure 2** shows that the radius depends almost linearly on a and reveals that the original value of r_E can be reproduced if a, determined in the first step of our analysis, is used, but with the opposite (wrong) sign.

To confirm this hypothesis, we again fitted model (2) to the data with $Q^2 < 1.05\,\text{fm}^{-2}$, but this time kept the radius fixed at 0.805(11) fm and adjusted only a. We obtained

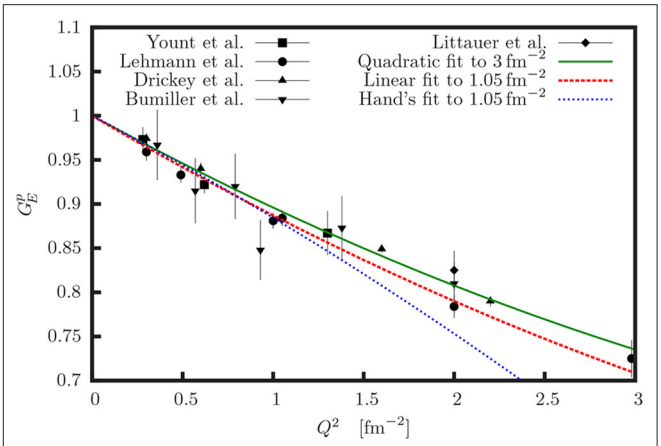


FIGURE 1 | The experimental data [23–27] considered in the analysis. The solid green line shows model (2) when both r_E and a are fitted to the data with $Q^2 \leq 3\,\text{fm}^{-2}$. The dashed red line shows the results when r_E is fitted to the data with $Q^2 \leq 1.05\,\text{fm}^{-2}$, while the parameter $a = 0.00787\,\text{fm}^4$ is kept constant. The blue dotted line corresponds to the original result of Hand et al. [2] assuming $a = -0.00787\,\text{fm}^4$.

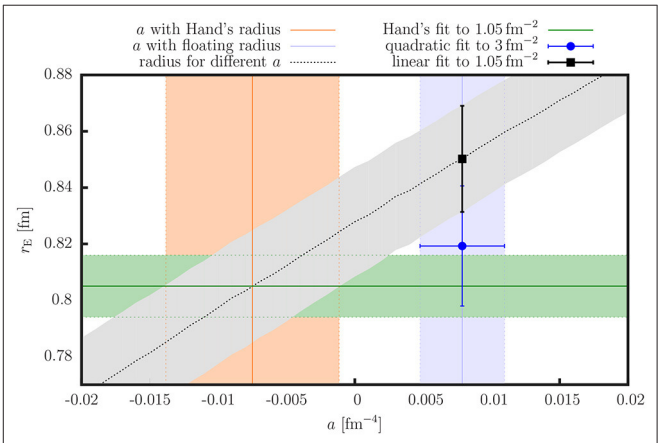


FIGURE 2 | The relation between parameters r_E and a that determine the model (2). The green band denotes the original result of Hand et al. [2]. The blue point represents the result of the analysis when both parameters are free and the model is fitted to the data with $Q^2 < 3.00\,\text{fm}^{-2}$. The vertical blue band indicates the value of the parameter a. The black point shows the final radius obtained by using the original two step approach of Hand et al. The gray line with the corresponding uncertainty shows how the extracted radius changes when a is modified from −0.02 to 0.02. The orange vertical band represents the result of the fit when only a is being fitted, while the radius is kept fixed at 0.805(11) fm. The cross-section of green, orange, and gray bands defines the area of possible values of a considered in the original analysis of Hand et al. [2]. The obtained result supports the hypothesis that a mistake has been made in the original analysis and that a was considered with the wrong sign.

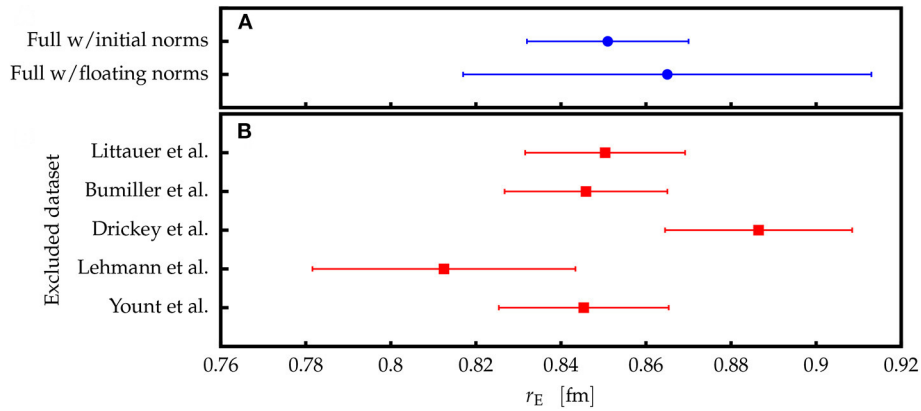


FIGURE 3 | The extracted values of the proton charge radius. **(A)** The difference between the value obtained with the fixed and floating normalization parameters. Addition of five free parameters significantly increases the uncertainty of the radius. **(B)** Calculated radii when performing the analysis with only four out of five data sets, demonstrating a tension between the data sets of Drickey et al. [25] and Lehmann et al. [27].

$a = -0.00749(63)\,\mathrm{fm}^4$, which strongly supports our assumption that a mistake was made in the original analysis. Additionally, our analysis has also revealed that the original study failed to acknowledge the uncertainty of a in the determination of r_E. Their analysis considered only statistical uncertainty and thus underestimated the final uncertainty of the radius.

To test the stability of the extracted radius, we have repeated the analysis by using all combinations of four of the five data sets. The results presented in **Figure 3** demonstrate the tension between the two most precise data sets, Drickey et al. [25] and Lehmann et al. [27]. The data of Lehmann et al. prefer a larger value of the proton charge radius and dominate the result when considering the data with small Q^2. The data of Drickey et al., on the other hand, favor a smaller proton charge radius and control the result at $Q^2 > 1.4\,\mathrm{fm}^{-2}$. While the discrepancy is too small to exclude a statistical fluctuation in the data, the most probable source of the tension are unaccounted for systematic effects, e.g., offsets in the absolute normalization of the reported data. The tension between the data is reduced if the normalizations of the data sets are kept as free parameters, as is being done in modern analyses of form factor measurements [15, 22, 28], but does not disappear completely. Furthermore, introduction of additional five free parameters to the fits (normalizations) increases the variance of the extracted result and dilutes the significance of the extracted radius, which in the given case equals to 0.865(48) fm (see **Figure 3**).

3. ROBUST ANALYSIS

The key problem of radius calculation is our ignorance of the true functional form of the proton charge form factor. Consequently, the form factor is approximated by various parameterizations. So far we considered function (2). Although the model was applied carefully to the data, it is not clear whether the quadratic function is an acceptable model for its description. The choice of a model can impact the result and can lead to a biased radius, i.e., a value that is systematically different from the true value. The bias is associated with the nature of the function and is typically smaller for functions with more free parameters. However, models with many parameters are justifiable only when data sets with large kinematic range and sufficient precision are available. Otherwise the variance of the radius increases to the level that the obtained result has no practical value. Hence, a model needs to be selected that exhibits a minimal bias of the extracted radius while keeping the variance of the result reasonably small. To achieve this, we have complemented the original analysis with a different technique based on a Monte-Carlo study of different form factor models, and are able to offer a more reliable determination of the radius.

Since the majority of the available data were measured only at small Q^2 and with limited precision, we investigated only models that depend on up to three parameters in order to keep the uncertainty of the extracted radius below the difference between the two competing values of the proton radius problem. Beside model (2), we considered:

$$G_{\mathrm{cubic}} = 1 + n_1 Q^2 + n_2 Q^4 + n_3 Q^6\,, \tag{3}$$

$$G_{\mathrm{Padé\ (0,1)}} = \frac{1}{1 + m_1 Q^2}\,, \tag{4}$$

$$G_{\mathrm{Padé\ (0,2)}} = \frac{1}{1 + m_1 Q^2 + m_2 Q^4}\,, \tag{5}$$

$$G_{\mathrm{hybrid}} = \frac{1 + n_1 Q^2 + n_2 Q^4}{1 + m_4 Q^8}\,, \tag{6}$$

$$G_{\mathrm{dipole}} = \frac{1}{(1 + m_1 Q^2)^2}\,, \tag{7}$$

where n_1, n_2, n_3, m_1, m_2, and m_4 represent adjustable parameters of the models. Using these parameters the r_E for each model can be calculated using Equation (1). The quadratic (Equation 2) and cubic functions (Equation 3) were considered as well as four rational functions. They are interesting because, like the dipole model, they introduce higher order terms and define the curvature of the form factor at higher Q^2, although they depend on relatively few parameters. For

completeness, we considered also the dipole model, which is known to report biased results [29], but can serve as a test of our approach.

The evaluation of the chosen models and tests of their capacity to reliably extract the radius can not be performed on the real data. Therefore, we developed a Monte-Carlo simulation

TABLE 2 | Summary of the Monte-Carlo study of the form-factor models (2) – (7).

Form factor	Simulation					Data	
model	Q^2_{best} [fm^{-2}]	Simulated bias [fm]	Simulated uncertainty [fm]	RMSE	Acceptable	Extracted radius [fm]	Standard error [fm]
Quadratic	2.9	−0.023	0.037	2.93	No	0.827	0.023
Cubic	5.4	−0.016	0.038	2.52	No	0.848	0.032
Padé (0, 1)	2.2	0.011	0.022	1.54	Yes	0.841	0.009
Padé (0, 2)	4.6	−0.015	0.028	2.09	No	0.826	0.026
Hybrid	5.1	−0.016	0.037	2.49	No	0.843	0.032
Dipole	1.5	−0.022	0.029	2.63	No	0.854	0.019

For every model listed in column one, the table shows the results for the most pessimistic case, as can be seen in ***Figure 4****. Column two shows the "best" value of* Q^2_{max} *at which* RMSE *reaches its minimum and defines the range of the data* $[0, Q^2_{best}]$ *to be used in the fit and in the extraction of the radius. Columns three and four contain the expected bias (extracted minus input radius) and uncertainty of the radius obtained with a chosen model. The best* RMSE *values for a specific model are presented in column five. A threshold for a good model is arbitrarily set at* $\sqrt{2}$*, see column six. The last two columns show the values of the proton charge radius extracted from the data, together with their standard errors.*

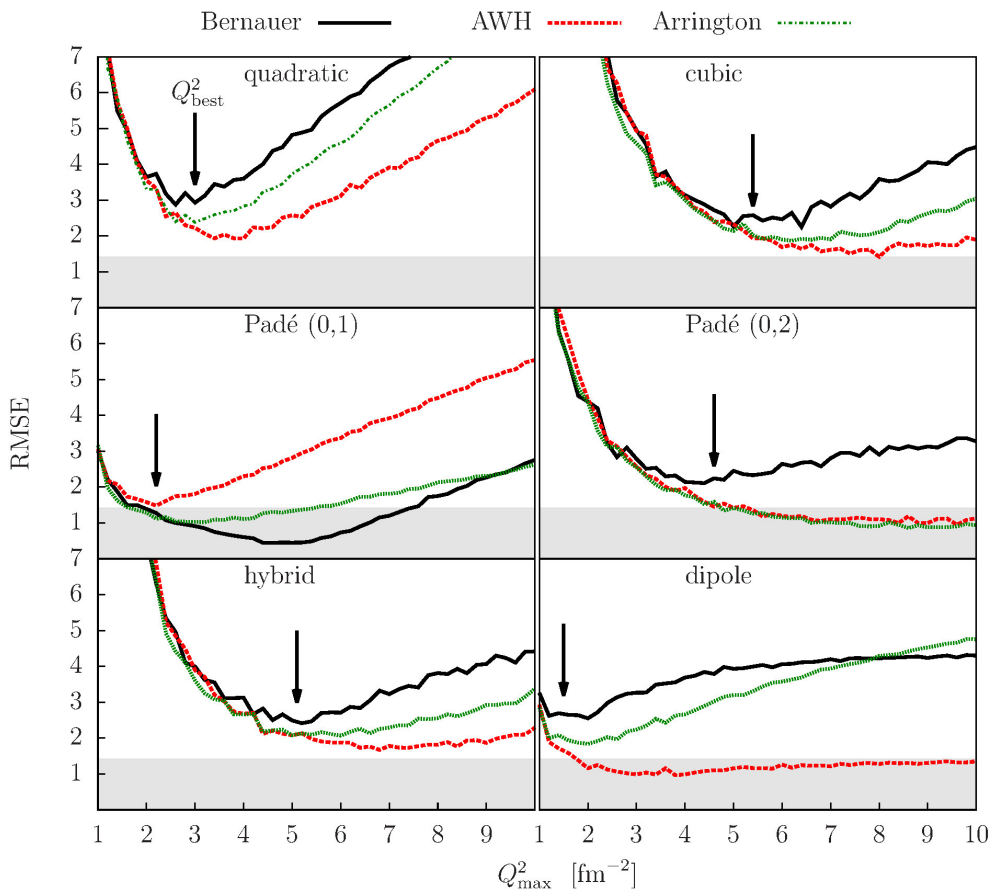


FIGURE 4 | Results of the Monte-Carlo study of the form-factor models (2)–(7). RMSE as a function of Q^2_{max}, obtained with realistic form factor parameterizations is used to evaluate the behavior of each model. According to our selection criterion a model is appropriate for the analysis if the minima of all the curves on a given plot lie below the threshold of $\approx \sqrt{2}$. The selection threshold is marked on the plots with gray bands. The black arrows on each plot denote the positions of the highest minimum which determines the interval $[0, Q^2_{best}]$ of the data that should be considered in the fit.

which generated many sets of pseudo data on a desirable kinematic interval using specific form factor models with known corresponding radii. These pseudo data were used to establish statistically relevant estimates on the size of the bias and variance of the extracted radius. The goal was to find a model that would (for a chosen kinematic range) return a radius with uncertainty smaller than $\sigma_{r_E} \leq \sigma_0 = 0.02\,\text{fm}$ and with the bias below $\Delta r_E \leq 1/(2\sigma_0)$. Therefore, we have defined the estimator

$$\text{RMSE} = \sqrt{\left(\frac{2\Delta r_E}{\sigma_0}\right)^2 + \left(\frac{\sigma_{r_E}}{\sigma_0}\right)^2} \tag{8}$$

which combines both conditions and could be used to quantify the quality of the selected model and search for the model with RMSE $\leq \sqrt{2}$. The six models were tested by using the parameterization of Bernauer et al. [22] determined from real data, the fifth-order continued-fraction model of Arrington and Sick [30], and the theoretical prediction of Alarcon et al. [20]. For each parameterization the pseudo data were generated and studied on the interval $[0, Q^2_{\text{max}}]$. The results of the analysis are gathered in **Table 2** and presented in **Figure 4**.

At small momentum transfers, the value of RMSE(Q^2_{max}) is governed by the variance, which decreases with the increasing number of data points considered in the fit. For large Q^2_{max}, the model is no longer capable of satisfactorily describing the data. Consequently, the extracted radius becomes biased and the RMSE(Q^2_{max}) again starts to increase. The position of the minimum determines the ideal momentum transfer range over which a given model gives the most reliable radius for a chosen form factor parameterization. Unfortunately, since we do not know the true functional form of the charge form factor, one cannot simply select a minimum from a single specific parameterization. Thus, we try to be conservative and choose the minimum with the highest RMSE value, Q^2_{best}, assuming that the form-factor parameterizations considered in the analysis form a

TABLE 3 | The parameters for the form-factor models (2), (4), (6), and (7), which have more than one free parameter.

Parameter	Extracted value	Relative significance
Quadratic		
r	0.827(23) fm	−1.30
a	6.0(24) fm^4	0.30
Padé (0, 2)		
m_1	2.92(18) fm^2	−0.94
m_2	1.7(25) fm^4	−0.14
Cubic		
n_1	−3.08(24) fm^2	−1.24
n_2	11.3(57) fm^4	0.95
n_3	−40.2(322) fm^6	−0.71
Hybrid		
n_1	−3.04(22) fm^2	−1.27
n_2	8.9(40) fm^4	0.74
m_4	275(236) fm^8	−0.63

Table shows the values for a given model extracted from the data. The relative contributions of the terms equipped with the given parameters to the total value of the form-factor at Q^2_{best} are also presented. The alternating signs of the parameters of the quadratic model (r, a) and cubic function (n_1, n_2, n_3) indicate that the true nature of the form-factor is more complex than a low order polynomial, thus requiring higher-order terms to match its slope and the curvature in a chosen Q^2-range. The positive values of m_1, m_2, and m_4 show that the Padé (0, 2) and the hybrid model do not have poles, while automatically ensure a correct asymptotic behavior of the form-factor. The large uncertainties of the higher-order terms (n_2, n_3, m_2, m_4) are governed by the large uncertainties of the available measurements.

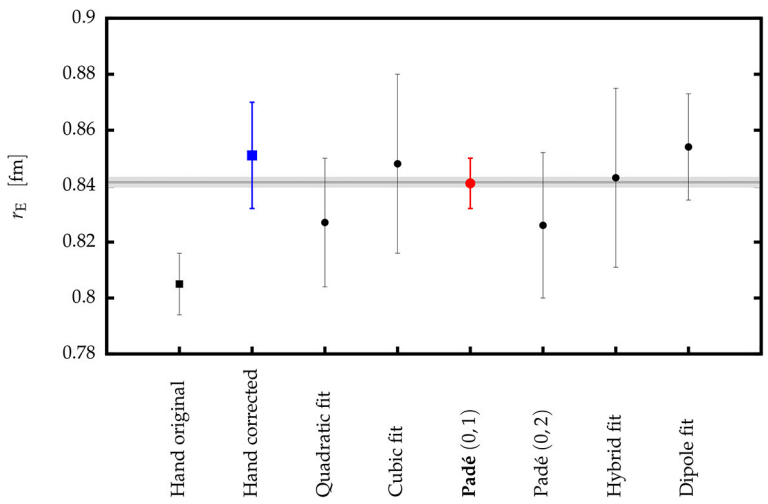


FIGURE 5 | The comparison of the extracted proton charge radii. The square points show the value calculated with the classical approach described in section 2 and the original result of Hand et al. [2]. The circles represent the model-dependent extractions of the radius obtained with the new analysis technique presented in section 3. The error bars show corresponding standard errors. According to the Monte-Carlo simulation the most robust estimate for the radius can be obtained using model (4), shown with the red circle. The gray band represents the new recommended value (CODATA 2018, [31]).

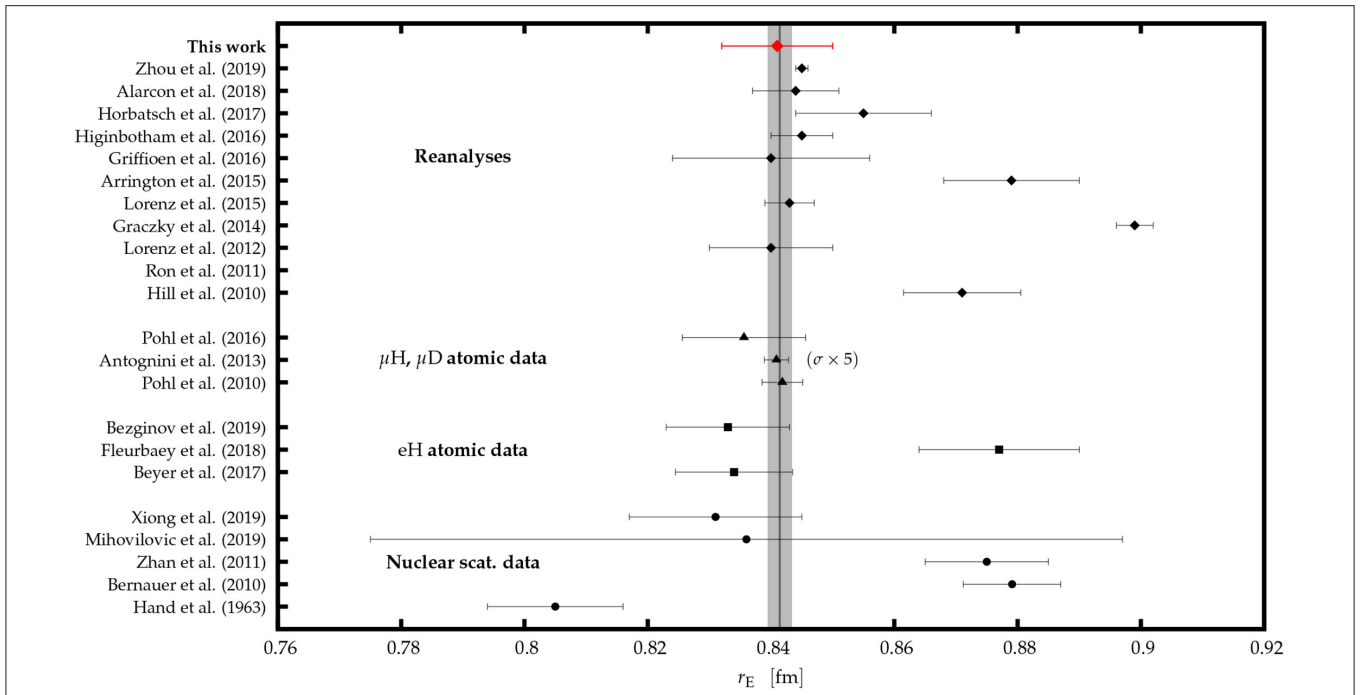


FIGURE 6 | The result of this work compared to other extractions of the proton charge radius. Full circles show findings of modern nuclear scattering experiments [11, 32–34] together with the original result of Hand et al. [2]. Full squares represent values obtained from the recent atomic hydrogen spectroscopy measurements [9, 10, 12]. The triangles denote values determined from the muonic hydrogen (deuterium) measurements [6, 7, 35]. The uncertainties of data from Pohl et al. and Antognini et al. are multiplied by factor 5 for clarity. The diamonds show recent reanalyses of the electron scattering experiments [14, 15, 17–20, 36–40]. The gray line with the corresponding band is the recommended value (CODATA 2018, [31]).

representative set of functions and that the true form factor may be somewhere in-between.

Once the Q^2_{best} for each of the models was estimated, the data could be fitted on the interval $[0, Q^2_{\text{best}}]$ and the proton charge radius could be determined. The results of the fits to the real data are shown in **Tables 2**, **3** and in **Figure 5**. However, the Monte-Carlo analysis demonstrates that only model (4) satisfies the condition for the RMSE $= 1.54 \approx \sqrt{2}$. All other models have RMSE values >2, which means that the radius results will not meet our criterion regarding the bias and variance. While quadratic and dipole functions are expected to have a large bias and should therefore be excluded, the remaining functions could still be considered, because their RMSE values are dominated by the large variance, but the calculated radii are expected to have large uncertainties. Hence, our best estimate for the radius is obtained with the Padé $(0, 1)$ approximant, yielding the radius of 0.841(9) fm.

4. CONCLUSIONS

In this paper we reanalyzed the proton charge form factor data from classical experiments performed in the 1960s by utilizing modern analysis tools that were not available at the time of the original analysis. Repeating the steps of Hand et al., we determined the radius to be 0.851(19) fm, a value which is 5 % larger than the result of the original paper. Using Monte-Carlo simulation we determined that the observed discrepancy is most probably related to a mistake in the interpretation of the Q^4-term when fitting the radius. To evaluate and minimize the dependence of the radius on the model applied in the analysis, the classical approach was superseded by a Monte Carlo-based analysis using pseudo-data generated with realistic form-factor parameterizations. In this approach the most appropriate fitting interval and the model function was selected by using a predefined selection criterion RMSE $\leq \sqrt{2}$. Among the considered functions only Padé $(0, 1)$ fulfilled the set condition. Using this function the best estimate for the proton charge radius was determined to be 0.841(9) fm. The obtained result is in good agreement with recent extractions of the radius and with the new recommended value (CODATA2018, [31]) (see **Figure 6**). Minimization of the model dependence of the extracted radius is key for reaching consistent interpretation of the modern electron scattering data. Here we offer an approach, which, relying on predefined selection criterion and using Monte-Carlo simulations, simultaneously examines both the model bias and variance. The method successfully applied to the data of Hand et al. can be directly extended to more complex models and used for a robust interpretation of the recent data.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://journals.aps.org/rmp/pdf/10.1103/RevModPhys.35.335.

AUTHOR CONTRIBUTIONS

MM, DH, MB, and SŠ have all provided substantial contributions to the analysis, interpretation of the data, and together drafted the paper. They all agree that the paper is ready for publication and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

This work was supported by the Federal State of Rhineland-Palatinate, by the Deutsche Forschungsgemeinschaft with the Collaborative Research Center 1044, by the Slovenian Research Agency under Grant P1-0102, and by Jefferson Science Associates which operates Jefferson Lab for the U.S. Department of Energy under contract DE-AC05-06OR23177.

REFERENCES

1. Miller GA. Defining the proton radius: a unified treatment. *Phys Rev C.* (2019) **99**:035202. doi: 10.1103/PhysRevC.99.035202
2. Hand LN, Miller DG, Wilson R. Electric and magnetic form factors of the nucleon. *Rev Mod Phys.* (1963) **35**:335–49. doi: 10.1103/RevModPhys.35.335
3. Hofstadter R, Bumiller F, Yearian MR. Electromagnetic structure of the proton and neutron. *Rev Mod Phys.* (1958) **30**:482–97. doi: 10.1103/RevModPhys.30.482
4. Weisenpacher P. Origin of the nucleon electromagnetic form-factors dipole formula. *Czech J Phys.* (2001) **51**:785–90. doi: 10.1023/A:1011618315454
5. Mohr PJ, Taylor BN, Newell DB. CODATA recommended values of the fundamental physical constants: 2010. *Rev Mod Phys.* (2012) **84**:1527–605. doi: 10.1103/RevModPhys.84.1527
6. Pohl R, Antognini A, Nez F, Amaro FA, Biraben F, Cardoso JMR, et al. The size of the proton. *Nature.* (2010) **466**:213–6. doi: 10.1038/nature09250
7. Antognini A, Nez F, Schuhmann K, Amaro FD, Birab F, Cardoso JMR, et al. Proton structure from the measurement of 2S-2P transition frequencies of muonic hydrogen. *Science.* (2013) **339**:417–20. doi: 10.1126/science.1230016
8. Pohl R, Gilman R, Miller GA, Pachucki K. Muonic hydrogen and the proton radius puzzle. *Annu Rev Nucl Part Sci.* (2013) **63**:175–204. doi: 10.1146/annurev-nucl-102212-170627
9. Beyer A, Maisenbacher L, Matveev A, Pohl R, Khabarova K, Grinin A, et al. The Rydberg constant and proton size from atomic hydrogen. *Science.* (2017) **358**:79–85. doi: 10.1126/science.aah6677
10. Bezginov N, Valdez T, Horbatsch M, Marsman A, Vutha AC, Hessels EA. A measurement of the atomic hydrogen Lamb shift and the proton charge radius. *Science.* (2019) **365**:1007–12. doi: 10.1126/science.aau7807
11. Xiong W, Gasparian A, Gao H, Dutta D, Khandaker M, Liyanage N, et al. A small proton charge radius from an electron-proton scattering experiment. *Nature.* (2019) **575**:147–50. doi: 10.1038/s41586-019-1721-2
12. Fleurbaey H, Galtier S, Thomas S, Bonnaud M, Julien L, Biraben F, et al. New measurement of the $1S - 3S$ Transition frequency of hydrogen: contribution to the proton charge radius puzzle. *Phys Rev Lett.* (2018) **120**:183001. doi: 10.1103/PhysRevLett.120.183001
13. Horbatsch M, Hessels EA. Evaluation of the strength of electron-proton scattering data for determining the proton charge radius. *Phys Rev C.* (2016) **93**:015204. doi: 10.1103/PhysRevC.93.015204
14. Griffioen K, Carlson C, Maddox S. Consistency of electron scattering data with a small proton radius. *Phys Rev C.* (2016) **93**:065207. doi: 10.1103/PhysRevC.93.065207
15. Higinbotham DW, Kabir AA, Lin V, Meekins D, Norum B, Sawatzky B. Proton radius from electron scattering data. *Phys Rev C.* (2016) **93**:055207. doi: 10.1103/PhysRevC.93.055207
16. Lee G, Arrington JR, Hill RJ. Extraction of the proton radius from electron-proton scattering data. *Phys Rev D.* (2015) **92**:013013. doi: 10.1103/PhysRevD.92.013013
17. Graczyk KM, Juszczak C. Proton radius from Bayesian inference. *Phys Rev C.* (2014) **90**:054334. doi: 10.1103/PhysRevC.90.054334
18. Lorenz IT, Meißner UlfG. Reduction of the proton radius discrepancy by 3σ. *Phys Lett B.* (2014) **737**:57–9. doi: 10.1016/j.physletb.2014.08.010
19. Horbatsch M, Hessels EA, Pineda A. Proton radius from electron-proton scattering and chiral perturbation theory. *Phys Rev C.* (2017) **95**:035203. doi: 10.1103/PhysRevC.95.035203
20. Alarcón JM, Higinbotham DW, Weiss C, Ye Z. Proton charge radius extraction from electron scattering data using dispersively improved chiral effective field theory. *Phys Rev C.* (2019) **99**:044303. doi: 10.1103/PhysRevC.99.044303
21. Bernauer JC, Achenbach P, Ayerbe Gayoso C, Böhm R, Bosnar D, Debenjak L, et al. High-precision determination of the electric and magnetic form factors of the proton. *Phys Rev Lett.* (2010) **105**:242001. doi: 10.1103/PhysRevLett.105.242001
22. Bernauer JC, Distler MO, Friedrich J, Walcher T, Ayerbe Gayoso PAC, Bühm R, et al. Electric and magnetic form factors of the proton. *Phys Rev C.* (2014) **90**:015206. doi: 10.1103/PhysRevC.90.015206
23. Littauer RM, Schopper HF, Wilson RR. Scattering of Bev electrons by hydrogen and deuterium. *Phys Rev Lett.* (1961) **7**:141–3.
24. Bumiller F, Croissiaux M, Dally E, Hofstadter R. Electromagnetic form factors of the proton. *Phys Rev.* (1961) **124**:1623–31.
25. Drickey DJ, Hand LN. Precise neutron and proton form factors at low momentum transfers. *Phys Rev Lett.* (1962) **9**:521–4. doi: 10.1103/PhysRevLett.9.521
26. Yount D, Pine J. Scattering of high-energy positrons from protons. *Phys Rev.* (1962) **128**:1842–9. doi: 10.1103/PhysRev.128.1842
27. Lehmann P, Taylor RE, Wilson R. Electron–proton scattering at low momentum energies. *Phys Rev.* (1962) **126**:1183.
28. Mihovilovič M, Weber AB, Achenbach P, Beranek T, Berič ič J, et al. First measurement of proton's charge form factor at very low Q^2 with initial state radiation. *Phys Lett B.* (2017) **771**:194–8. doi: 10.1016/j.physletb.2017.05.031
29. Bernauer JC, Distler MO. Avoiding common pitfalls and misconceptions in extractions of the proton radius. In: *ECT* Workshop on The Proton Radius Puzzle*. Trento (2016).
30. Arrington J, Sick I. Precise determination of low-Q nucleon electromagnetic form factors and their impact on parity-violating *e-p* elastic scattering. *Phys Rev C.* (2007) **76**:035201. doi: 10.1103/PhysRevC.76.035201
31. Tiesinga E, Mohr PJ, Taylor BN, Newell DB. *The 2018 CODATA Recommended Values of the Fundamental Physical Constants* (2019). Available online at: http://physics.nist.gov/constants
32. Bernauer JC. *Measurement of the Elastic Electron-Proton Cross Section and Separation of the Electric and Magnetic Form Factor in the Q^2 Range From 0.004 to 1 $(GeV/c)^2$*. Mainz U., Inst. Kernphys. (2010). Available online at: http://wwwa1.kph.uni-mainz.de/A1/publications/doctor/bernauer.pdf
33. Zhan X, Allada K, Armstrong DS, Arrington JR, Bertozzi W, Boeglin W, et al. High-precision measurement of the proton elastic form factor ratio $\mu_p G_E/G_M$ at low Q^2. *Phys Lett B.* (2011) **705**:59–64. doi: 10.1016/j.physletb.2011.10.002
34. Mihovilovič M, Merkel H. ISR experiment at A1-Collaboration. *EPJ Web Conf.* (2019) 218:04001. doi: 10.1051/epjconf/201921804001
35. Pohl R. Laser spectroscopy of muonic hydrogen and the puzzling proton. *J Phys Soc Jpn.* (2016) **85**:091003. doi: 10.7566/JPSJ.85.091003
36. Arrington J, Sick I. Evaluation of the proton charge radius from E-P scattering. *J Phys Chem Ref Data.* (2015) **44**:031204. doi: 10.1063/1.4921430
37. Hill RJ, Paz G. Model independent extraction of the proton charge radius from electron scattering. *Phys Rev D.* (2010) **82**:113005. doi: 10.1103/PhysRevD.82.113005

38. Ron G, Zhan X, Glister J, Lee B, Allada K, Armstrong W, et al. Low Q^2 measurements of the proton form factor ratio $\mu_p G_E/G_M$. *Phys Rev C*. (2011) **84**:055204. doi: 10.1103/PhysRevC.84.055204
39. Lorenz IT, Hammer HW, Meißner UlfG. The size of the proton–closing in on the radius puzzle. *Eur Phys J A*. (2012) **48**:151. doi: 10.1140/epja/i2012-12151-1
40. Zhou S, Giuliani P, Piekarewicz J, Bhattacharya A, Pati D. Reexamining the proton-radius problem using constrained Gaussian processes. *Phys Rev C*. (2019) 99:055202. doi: 10.1103/PhysRevC.99.055202

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

OPTICS AND PHOTONICS

Alessandro Pitanti received his B.S. (2004) and M.S (2006) in physics from University of Pisa, and the Ph.D. (2010) in physics from University of Trento. After a postdoc at Scuola Normale Superiore in Pisa, in 2012 he joined the California Institute of Technology as a Marie-Curie fellow. Back in Italy, in 2014 he joined the NEST Lab. in Pisa, which is joint enterprise of Scuola Normale Superiore, CNR and IIT. Finally in 2017 he was promoted to be a permanent CNR researcher. His current interests include NIR and THz opto- and electro-mechanics and strain engineering of 2D materials. He leads the micro and nanomechanics initiatives at NEST. For his research activity, in 2015 he was awarded the Di Braccio prize by the Accademia dei Lincei. He has more than 100 papers, which collected more than 1200 citations.

The control of electromagnetic wave characteristics (intensity, phase, polarization) is becoming the cornerstone of light-based technologies, which find broad and diverse applications in areas ranging from telecommunications to high-precision measurements and quantum/classical computation. Successful approaches to light control originate from the field of artificial materials, where the first results in reproducing the optical properties of natural materials through nano-structuration quickly led to their recognition as powerful elements for photonic applications. When integrated with mechanical elements, metamaterials gain an extra dynamical dimension, which allows quickly changing the state of light, arbitrarily controlling intensity, phase or polarization. In this paper, it is demonstrated a wide range control of the polarization state of light in non-trivial path on the Poincaré sphere through mechanical actuation of a chiral dielectric metasurface. Exploiting the overdamped fundamental mechanical drum mode, a broad polarization modulation from 0.1 up to 1.4 MHz is achieved. With the potential of a further frequency increase through the use of high-order mechanical modes, this device represents a novel element for chip-scale light control and modulation.

frontiers in Physics

ORIGINAL RESEARCH
published: 10 January 2020
doi: 10.3389/fphy.2019.00231

Broadband Dynamic Polarization Conversion in Optomechanical Metasurfaces

Simone Zanotto[1], *Martin Colombano*[1,2], *Daniel Navarro-Urrios*[3], *Giorgio Biasiol*[4], *Clivia M. Sotomayor-Torres*[2,5], *A. Tredicucci*[1,6] *and Alessandro Pitanti*[1]*

[1] NEST Lab., National Research Council—Istituto Nanoscienze and Scuola Normale Superiore, Pisa, Italy, [2] Catalan Institute of Nanoscience and Nanotechnology (ICN2), Spanish National Research Council and Barcelona Institute of Science and Technology, Bellaterra, Spain, [3] MIND, Institute of Nanoscience and Nanotechnology of the University of Barcelona, Departament d'Electrònica, Facultat de Física, Universitat de Barcelona, Barcelona, Spain, [4] Istituto Officina dei Materiali National Research Council, Laboratorio TASC, Basovizza, Italy, [5] ICREA—Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain, [6] Dipartimento di Fisica, Università di Pisa, Pisa, Italy

Artificial photonic materials, nanofabricated through wavelength-scale engineering, have shown astounding and promising results in harnessing, tuning, and shaping photonic beams. Metamaterials have proven to be often outperforming the natural materials they take inspiration from. In particular, metallic chiral metasurfaces have demonstrated large circular and linear dichroism of light which can be used, for example, for probing different enantiomers of biological molecules. Moreover, the precise control, through designs on demand, of the output polarization state of light impinging on a metasurface, makes this kind of structures particularly relevant for polarization-based telecommunication protocols. The reduced scale of the metasurfaces makes them also appealing for integration with nanomechanical elements, adding new dynamical features to their otherwise static or quasi-static polarization properties. To this end we designed, fabricated and characterized an all-dielectric metasurface on a suspended nanomembrane. Actuating the membrane mechanical motion, we show how the metasurface reflectance response can be modified, according to the spectral region of operation, with a corresponding intensity modulation or polarization conversion. The broad mechanical resonance at atmospheric pressure, centered at about 400 kHz, makes the metasurfaces structure suitable for high-frequency operation, mainly limited by the piezo-actuator controlling the mechanical displacement, which in our experiment reached modulation frequencies exceeding 1.3 MHz.

Keywords: metasurface, optomechanics, polarization, chirality, nanomechanics

OPEN ACCESS

Edited by:
Lorenzo Pavesi,
University of Trento, Italy

Reviewed by:
Andrey Miroshnichenko,
University of New South Wales, Australia
Venu Gopal Achanta,
Tata Institute of Fundamental Research, India

***Correspondence:**
Alessandro Pitanti
alessandro.pitanti@nano.cnr.it

Specialty section:
This article was submitted to Optics and Photonics, a section of the journal Frontiers in Physics

Received: *31 October 2019*
Accepted: *10 December 2019*
Published: *10 January 2020*

Citation:
Zanotto S, Colombano M, Navarro-Urrios D, Biasiol G, Sotomayor-Torres CM, Tredicucci A and Pitanti A (2020) Broadband Dynamic Polarization Conversion in Optomechanical Metasurfaces. Front. Phys. 7:231. doi: 10.3389/fphy.2019.00231

INTRODUCTION

Light control through nanoengineered materials has recently risen as one of the core business of photonics [1–3]. The great degree of flexibility with which an artificial dielectric function can be designed, paired with powerful optimization tools [4, 5] made the use of artificial photonic materials widely available, with many technological applications currently reaching the market [6, 7]. A common implementation of such structures takes the form of engineered surfaces (metasurfaces), where the device planar size is larger than its thickness, making them compatible with standard nanofabrication techniques as well as easy to integrate on chip. Most

of the metasurfaces operate in a static configuration: given an input in terms of an impinging monochromatic light beam, they are able to output another beam, in transmission or reflection, with controllable wavevector, amplitude, phase, and polarization. A highly desirable feature would be the possibility to tune and control the output beam by acting on the metasurface, in such a way to obtain modulation, switching or more complex functionalities. Few implementations have tried to achieve this goal, showing the possibility to statically reconfigure the metasurface geometry in such a way to obtain multi-state, static response from a single device: the tuning mechanism includes the use of electrostatic forces [8, 9], stretchable substrates [10], static optical forces [11], or phase-changing materials [12]. More recently, several groups have used temperature change as the main tuning mechanism [13, 14] reaching large frequency tunability. This approach has been used to create tunable notch-filters [15], and image [16], and polarization manipulation [17], albeit at quasi-static operation frequency. The venues opened by the recent field of cavity optomechanics have shown how micro- and nano-mechanical object can be successfully coupled with nanophotonics devices [18]. A particularly successful strategy sees the inclusions of electrical elements in optomechanical systems [19], either for a proper action-back action coupling or for a coherent mechanical excitation through shaking, electrostatic actuation, etc.

In this article, we employ an electro-optomechanical approach to modulate the metasurface response in time. By embedding a periodic pattern on a suspended dielectric membrane we show how the photonic response function can be modulated by mechanically actuating the membrane fundamental drum motional mode. In particular, by using a chiral pattern for the metasurface definition, we show a dynamical manipulation of the output light polarization, which can be dynamically controlled along non-trivial paths on the Poincarè sphere. A detailed discussion of quantitave, single frequency polarization modulation at ∼400 kHz and, conversely, light polarimetry has been reported elsewhere [20]; in this article we show how the operation bandwidth of the device can be extended to about 1.4 MHz, exploiting the overdamping regime of the mechanical resonator at atmospheric pressure.

DEVICE DESIGN AND FABRICATION

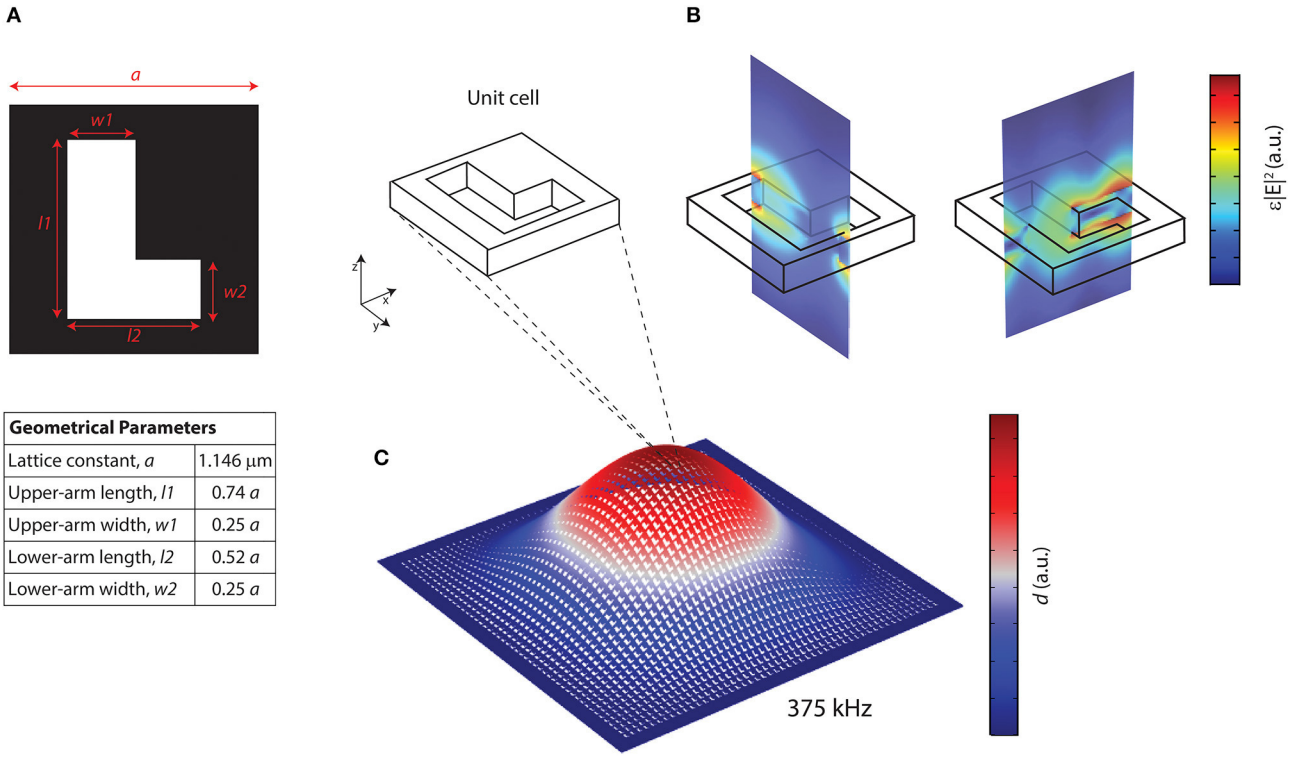
The metasurface has been devised considering a *minimal design* approach, by using patterned holes on a single dielectric slab. This positively compares to the metallic-pattern metasurfaces, which usually show a larger amount of ohmic losses, especially when operated in the near infrared range. Numerical simulations have been performed using the Periodic Patterned Multi-Layer (PPML) Matlab script, based on Rigorous Coupled Wave Analysis method (RCWA) code[1] For the basic hole shape, two orthogonal, joined rectangles have been considered, arranged as a "L"; this is one of the easiest shape to produce a chiral response and can be easily parametrized by considering the two arms lengths and widths, respectively, *l1*/*l2* and *w1*/*w2* as reported in the black and white sketch of **Figure 1A**, where black color indicates GaAs and white color air trenches. Starting with a 220 nm GaAs substrate, we performed a 4-parameter optimization in such a way to maximize the metasurface circular dichroism at 1,550 nm. Details on the optimization and a typical device photonic characterization has been reported elsewhere [21]. The final geometric parameters resulting from the optimization are reported in the table in **Figure 1**. The metasurface has been designed considering a periodic lattice *a* resulting in photonic modes which are delocalized across the whole structure: the single cell electrical energy density, $\varepsilon|E|^2$, with dielectric constant *e* and electric field *E*, obtained from FEM simulations for an infinite square lattice in the Γ point can be seen in **Figure 1B**. The simulated photonic modes can be reasonably compared with what we expect in the central region of the metasurface, given the large number of periods we fabricated (50 × 50 lattice constants).

The simulation of the low-frequency mechanical modes requires the geometry of the full suspended structure. Continuous mechanics FEM simulations have been used in order to find the resonant frequency of the first eigenmodes. GaAs has been parametrized using the elasticity matrix and material density [22], the former being properly rotated for our crystal orientation. An illustration of the fundamental drum mode is reported in **Figure 1C**; here the colormap indicates the modulus of the mechanical displacement, $d = \sqrt{u^2 + v^2 + w^2}$, with *u*, *v*, and *w* displacement components along *x*, *y*, and *z* directions, respectively, while the artificial displacement has been artificially exaggerated for visualization purposes. The mode resonant frequency has been found to be 372 kHz, in good agreement with independent characterization with a Laser Doppler Vibrometer [20].

The metasurface was fabricated starting with a GaAs wafer on which a 1.5 μm $Al_{0.5}GaAs_{0.5}$ sacrificial layer has been grown by Molecular Beam Epitaxy (MBE). On top of that, a further 220 nm GaAs (device) layer was further grown. An electron beam resist mask (AllResist AR-P6200) was spun on the sample and subsequently exposed using a 30 keV electronic beam. In particular, the single metasurface pattern has been replicated in such a way to cover roughly 50 × 50 μm areas (50 × 50 lattice constants) which also define the membrane size. Each area has been exposed using a different geometrical scaling factor (up to 10% modification) in order to slightly tune the photonic resonances. Among all the different fabricated metasurfaces we focused on six devices, L1–L6, which have been exposed considering the scaling factor reported in **Table 1**.

The exposed pattern has been transferred to GaAs using a Chlorine-based Inductively Coupled Plasma—Reactive Ion Etching (ICP-RIE, Sentech SI 500) machine; plasma has been ignited from a gas mixture of $BCl_3/Cl_2/Ar$, with 6/1/10 sccm, respectively. As a last step, the membranes were released in a pure HF solution. The fabricated devices have been preliminarily inspected with Scanning Electron Microscopy: a full schematic of the fabrication flow and a micrograph of a typical device are reported in **Figure 2**.

[1] Available online at: https://it.mathworks.com/matlabcentral/fileexchange/55401-ppml-periodically-patterned-multi-layer



Geometrical Parameters	
Lattice constant, *a*	1.146 μm
Upper-arm length, *l1*	0.74 *a*
Upper-arm width, *w1*	0.25 *a*
Lower-arm length, *l2*	0.52 *a*
Lower-arm width, *w2*	0.25 *a*

FIGURE 1 | (A) Standard geometrical parameters for the L-shaped unit cell. **(B)** FEM simulated electrical energy density for a typical metasurface mode. For improved visibility cut planes *xz* and *yz* are shown. **(C)** Modulus of the total displacement of the FEM simulated fundamental mechanical mode.

TABLE 1 | Scaling of the geometrical parameters reported in **Figure 1**.

Device number	Geometrical scaling factor
L1	1.02
L2	1
L3	0.99
L4	0.98
L5	0.96
L6	0.94

Only parameters l1, w1, l2, and w2 are scaled; the lattice constant a is always equal to 1.146 μm.

STATIC CHARACTERIZATION

The metasurface was characterized by reflectance spectra. Light coming from a tunable near-infrared laser (Newport TLD600) was shined on the device using a lens with 7 cm focal length. This produced a roughly 50 μm wide beam spot on the device layer. All experiments were performed at normal incidence; after interacting with the metasurface, the reflected laser light was sent into a detection stage through a beam splitter. The output signal was then detected either using a commercial polarimeter (Thorlabs PAX-1000, maximum sampling rate 600 Hz) or using a fast InGaAs detector (Newport, 1623). In the latter case, it is possible to add a linear polarizer filter to partially inspect the polarization state of the reflected light. A sketch of the experimental setup is shown in **Figure 3A**. The piezoelectric actuator and the lock-in amplifier (LIA) are used for the dynamic characterization and are initially disconnected.

At first, we have characterized a full set of metasurfaces which have been fabricated in the same run with different geometrical scaling factors (L1–L6). Impinging laser light polarization was linear horizontal (x-direction), in a wavelength range from 1,520 to 1,570 nm; no analyzer was present in front of the fast detector. The reflectivity spectra of different membranes are reported in **Figure 4A** and mostly show fast signal oscillations which originate from multiple reflections from the substrate, both considering the first air-GaAs interface as well as the bottom GaAs-air interface. These oscillations are modulated by slowly-varying envelopes coming from the proper metasurface resonances. This can be better seen by inserting the linear polarizer in front of the detector and rotate it in such a way to project the polarization state along the vertical linear polarization state (y-direction). This makes the detection system sensitive to polarization rotation. In particular, the metasurface resonances emerge, being the only physical system in the experimental line capable of modifying the polarization state of light. The cross-polarization reflectivity results are reported in **Figure 4B**, where the signal envelopes now take the shape of Fano-shaped [23] broad peaks, directly coming from the metasurface resonance. As can be seen, the small change of the geometrical parameters gives a shift in the metasurface resonances. In particular, device

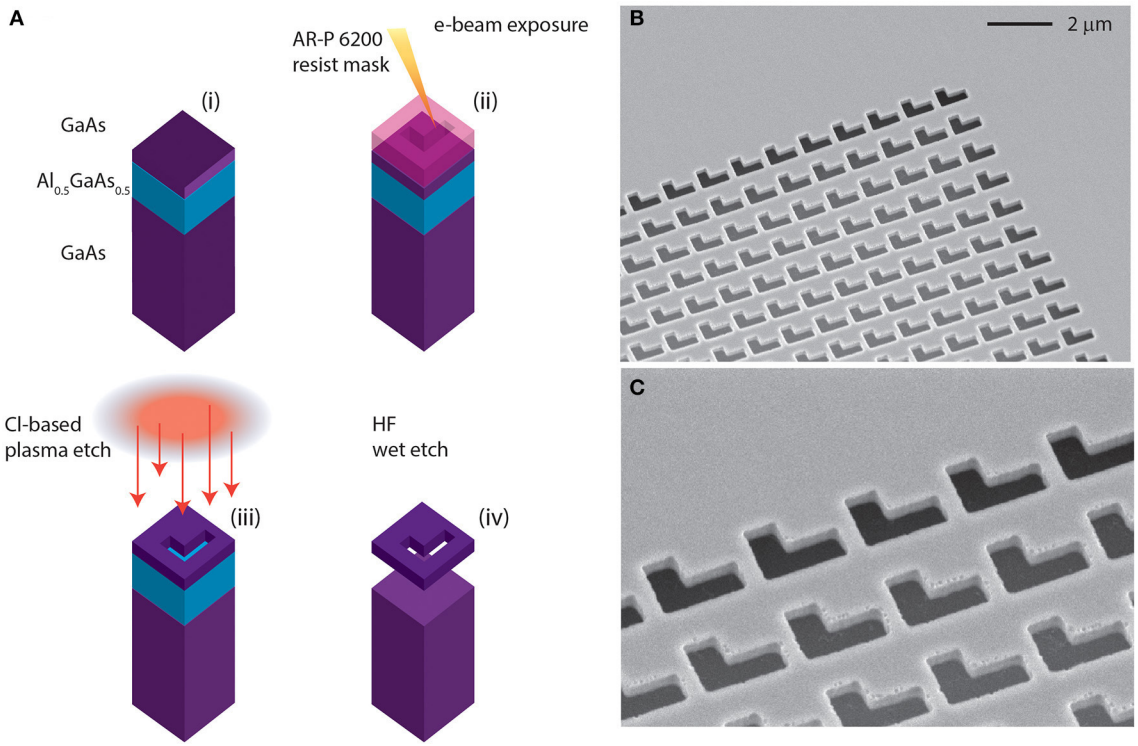


FIGURE 2 | (A) Fabrication steps. Starting from a GaAs/$Al_{0.5}GaAs_{0.5}$ heterostructure (i), we spin-coat a layer of AR-P 6200 resist (ii). E-beam lithography is used to define the patterns, (iii) which are transferred on the device layer through a dry etching step (iv). Finally, the full structure gets released through HF wet etching. The SEM micrograph **(B)** shows a typical fabricated device with a magnification of the pattern **(C)**.

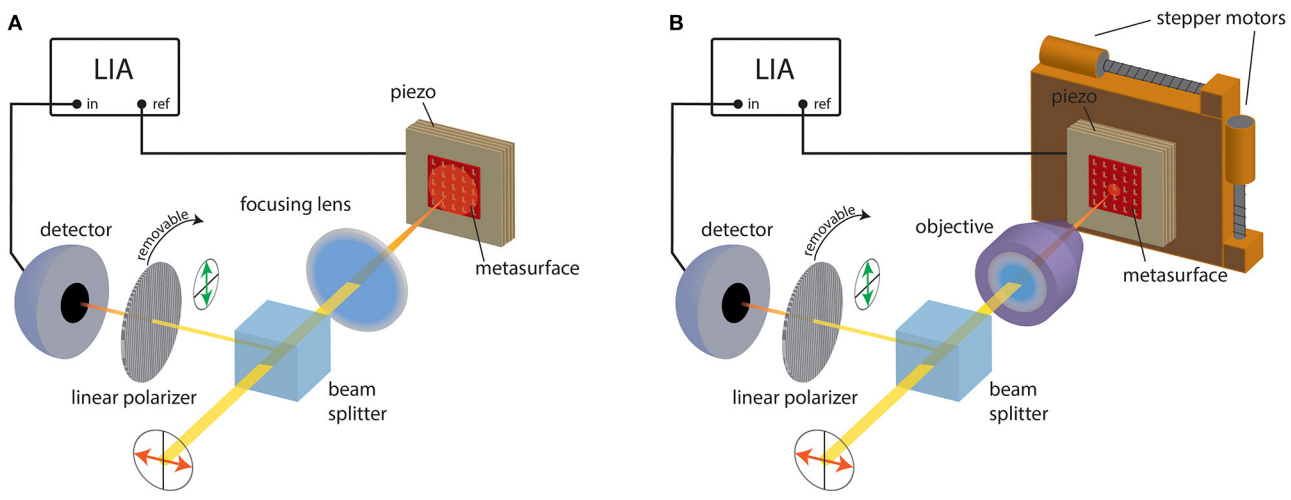


FIGURE 3 | Experimental set configurations. **(A)** Sketch of the experimental setup. The laser light is focused on the metasurface; the reflected signal is then sent to an IR detector through a beam splitter. A linear polarizer filter can be inserted before detection. The signal is then analyzed with a lock-in amplifier (LIA) whose reference signal is used to control the piezoelectric driving stack. **(B)** Same as **(A)** with a microscope objective instead of the focusing lens and the metasurface additionally sitting on a motorized stage for spatial maps. Note that in the sketches the laser beam size is not in scale.

L1 showed a localized, strong peak around 1,545 nm, making it the perfect candidate for a deeper photonic and mechanical characterization. To this end, we mounted the device in a slightly different setup, where the sample sits on a motorized stage and the laser light is focused through a 50×, long working distance objective (see **Figure 3B**). The stronger focusing effect

results in an estimated beam size of about 5 μm on the sample surface enabling local probing of the membrane. **Figure 5A** reports a map of the cross-polarized reflectivity at 1,532.5 nm around L1 central position. A distinct region representing the metasurface is clearly visible, owing to the device polarization rotation characteristic, whereas the contribution coming from the substrate is negligible thanks to the polarization filter effect. The full cross-polarized spectra of few selected points on the map have also been reported in **Figure 5**: as can be seen, the spectra recorded in the central region of the membrane all share a qualitatively similar shape, with small differences due to some inhomogeneity present on the membrane; note that the spectrum is quite different from the one shown in **Figure 4B**, due to the different excitation condition in terms of accessible wavevectors. The spectra recorded in the outer region show a residual, small polarization rotation effect, likely given by a small overlap of the laser beam with the outermost metasurface pattern.

To gain a better insight on the polarization conversion effect, we focused the laser light on point six of the membrane, at the same time changing the fast detector with the polarimeter. In this way, it is possible to follow the spectral evolution of the polarization state on the Poincaré sphere [see for example [24]], which is the standard tool used to illustrate the Stokes parameters S_1, S_2, and S_3 directly measured by the polarimeter [25]. The experimental results are shown in **Figure 5B**, where the colormap allows to associate points on the Poincaré sphere with the cross-polarized spectrum in panel number six. Far from the resonance, the polarization is almost equal to the input one (linear horizontal). By getting closer to the resonance, the polarization state is rotated becoming elliptical. Interestingly, it can be clearly seen that the state evolves in spiral loops, corresponding to the multiple Fabry-Perot resonances originated by the substrate. This is a clear indication that the overall phase shift given by the multiple reflection of light in the substrate strongly influences the metasurface operation and polarization conversion power. Within a single Fabry-Perot fringe, the phase of light undergoes a 2π shift. If the metasurface is influenced by the phase, one should expect a closed path on the Poincaré sphere when one single fringe is probed with the laser. The fact that the loops are not closed is due to the dispersion characteristic: in fact, a full phase round trip is concluded at a different wavelength than the one in the initial point of the sweep. Therefore, even if the same Fabry-Perot induced phase shift occurs, the different metasurface response at different wavelength results in a different conversion effect, clearly visible on the Poincaré sphere. This important observation points toward the strong possibility of changing the polarization conversion effect by simply acting on the phase of light reflected from the substrate; this can be easily done by changing the metasurface—substrate separation distance.

COHERENT DYNAMIC CHARACTERIZATION

The effect of mechanical modes on the metasurface polarization conversion effect can be evaluated by placing the sample on a piezoelectric multilayer stack, driven by a high-frequency local oscillator. This can be taken as a reference for demodulating the fast detector signal with a lock-in amplifier (LIA) (Zurich Instruments, UHF-LI). The signal read by the LIA is a direct measurement of the effect of the mechanical force on the state of light: by inserting/removing the analyzer, we can infer the case of general intensity modulation with the more relevant polarization conversion. In fact, if the polarization is not modified, the reflectivity spectra with and without the analyzer will be linearly proportional, with a scaling factor coming from the projection of the output (constant) polarization state on the linear vertical state.

Initially, we used the objective setup (**Figure 3B**) to focus the laser in the center of the membrane. Setting a laser wavelength of 1,532.5 nm, corresponding to a local maximum in one of the fringes of **Figure 4B**, we swept the piezo drive frequency while at the same time recording the LIA

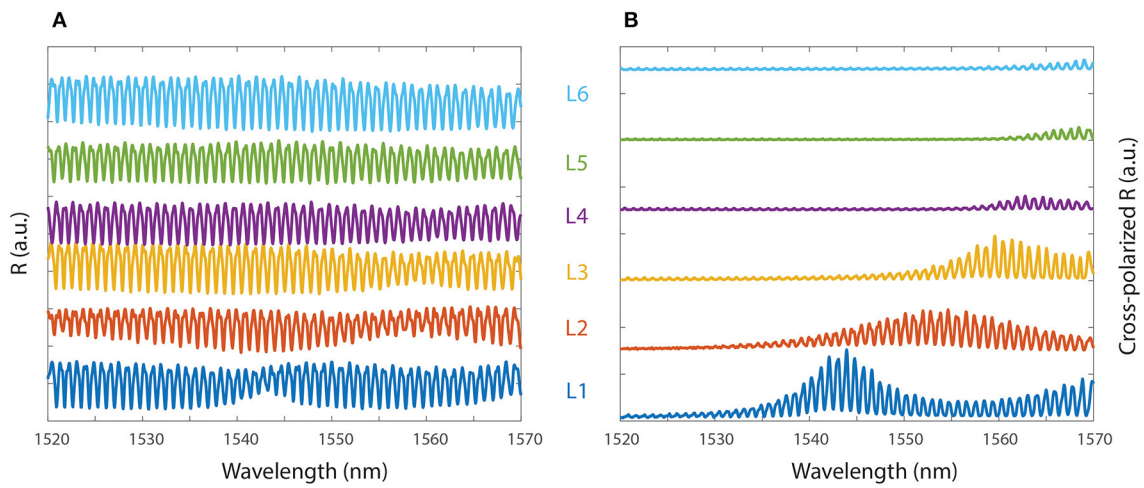


FIGURE 4 | Direct **(A)** and Cross-polarized **(B)** reflectivity spectra for different metasurfaces.

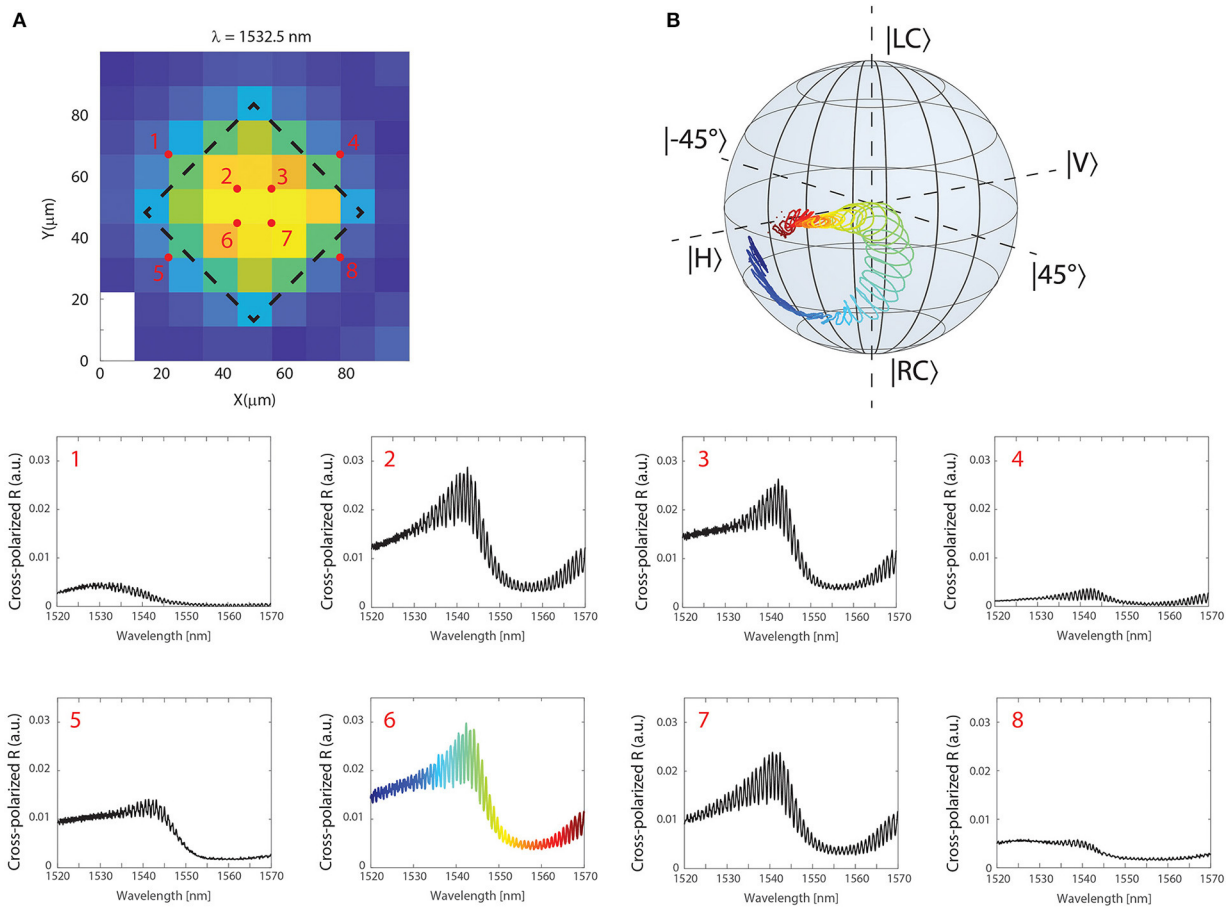


FIGURE 5 | (A) Map of the cross-polarized signal at 1,532.5 nm across the membrane, whose contour is indicated as a guide for the eyes. The spectra in selected points (1–8) are individually reported. **(B)** Spectral evolution on the Poincaré sphere probing the point 6 of the membrane. The cross-polarized spectrum and the Poincaré plot are correlated through the plot color.

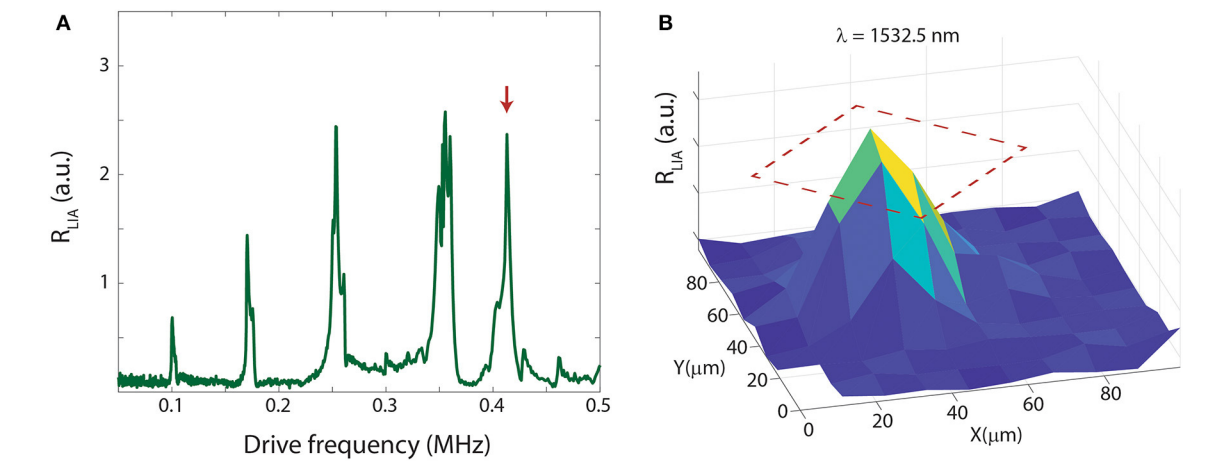


FIGURE 6 | (A) Piezo-driven mechanical spectrum at atmospheric pressure with the laser focused in the membrane center. The peaks identify the out-of-plane mechanical eigenmodes of the piezo stack. **(B)** Map of the LIA amplitude at 430 kHz [red arrow in **(A)**] around the membrane position. The full metasurface position is overlaid with a red-dashed line as a guide for the eyes.

amplitude signal, R_{LIA}. The result reported in **Figure 6A** shows a rich spectrum, which originates from the convolution of the piezoelectric stack out-of-plane modes and the membrane fundamental drum mode. This mode is strongly overdamped at atmospheric pressure and can be excited over a wide frequency range around its central resonance at about 413 kHz. Setting this frequency as a monochromatic drive of the piezo actuator, we probed the spatial dependence of the mechanical-induced polarization conversion around the membrane central position. The resulting map, reported in **Figure 6B**, is in good agreement with the expected mechanical simulations shown in **Figure 1C**.

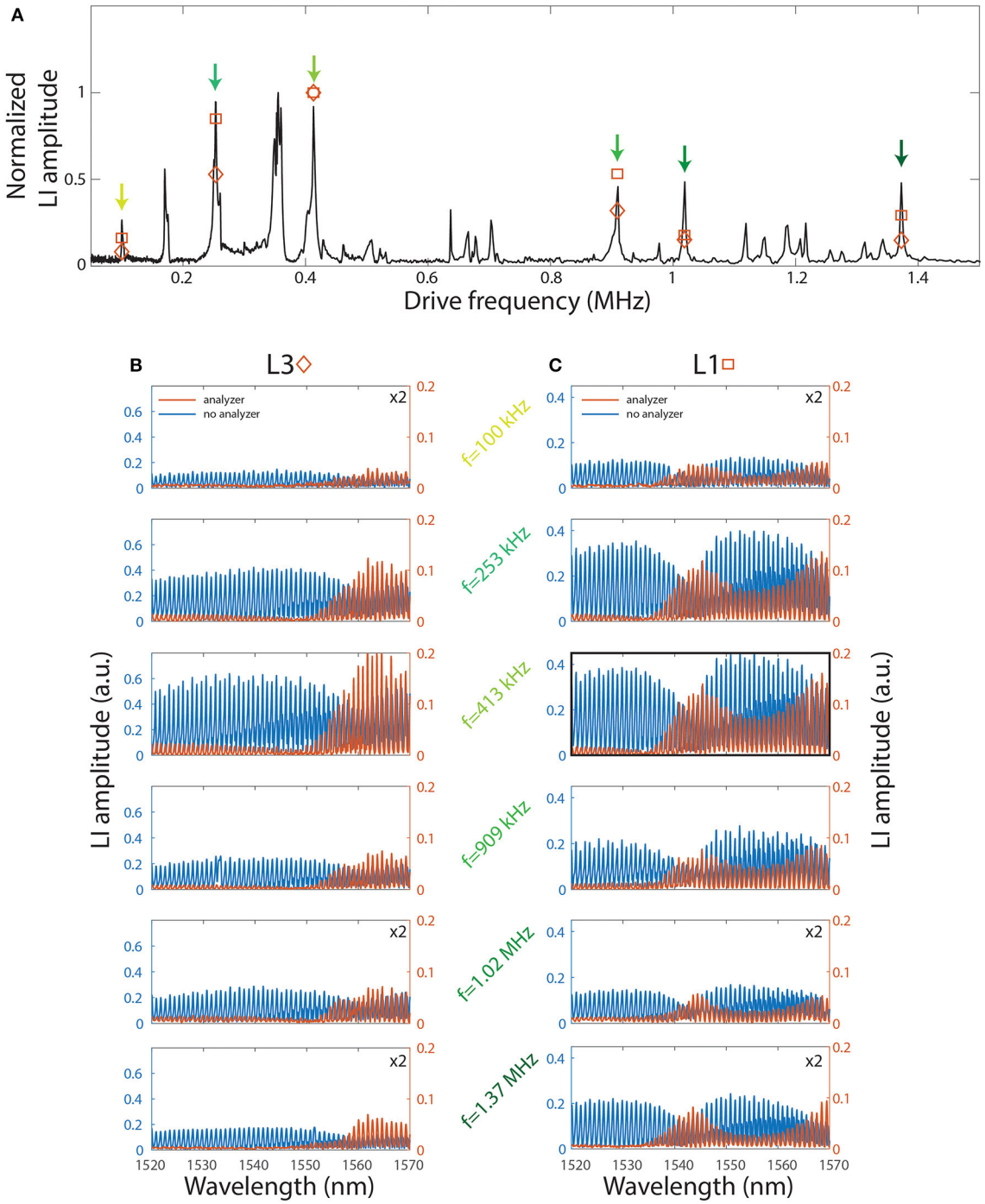


FIGURE 7 | (A) Broadband mechanical spectrum. LIA-demodulated spectra for several driving frequencies are reported for sample L3 **(B)** and L1 **(C)**.

Switching from the objective to the focusing lens, in order to avoid any effect due to possible local inhomogeneity of the sample, we scan a laser while keeping the piezo drive at 413 kHz with a constant amplitude of 100 mV. At first, we analyzed the direct mechanical-induced light intensity modulation; the measured R_{LIA} is reported as a blue trace in one of the inset of **Figure 7** (second column, third row, bold box), showing a non-negligible modulation effect even at wavelength far from the metasurface resonance, probably due to small changes in reflectivity of the whole chip with respect to its static configuration. On the other hand, upon the insertion of the analyzer, the spectral shape of the signal is strongly modified, suggesting that an intensity modulation can be accompanied by a pure polarization conversion effect (see **Figure 7**, same panel, red trace). Note that the slight mismatch of the signal fringes can be imputed by the different polarization components detected in the experiment. The largest conversion effect is found around the maximum of the cross-polarized signal reported in **Figure 4**. A quantitative calibration of the conversion effect, which is dynamically following non-trivial paths on the Poincaré sphere, can be obtained after a careful calibration of the system, which is out of the scope of this paper and can be found elsewhere [20].

The overdamped drum mode we are considering allows one to work at several different mechanical frequencies, mainly limited by the range of operation of the driving stack. To show that, we recorded a broad band mechanical spectrum, reported in **Figure 7A**, where narrow and strong peaks can be seen up almost 1.5 MHz. We selected several driving peaks for testing the metasurface polarization conversion effect, in a range from 100 kHz to ∼1.4 MHz (colored arrows in **Figure 7**). The results for devices L1 and L3 are reported in **Figures 7B,C**, respectively. First observation is that the spectral shapes of the signals R_{LIA} with (red curve) and without (blue curve) the polarization analyzer are significantly different at all the measured mechanical frequencies. Furthermore, for each spectrum, we can take the maximum of cross polarized R_{LIA}, which is the LIA amplitude signal when the analyzer is present. The results are superimposed in the broad mechanical spectrum of **Figure 7A**, showing a good correlation between mechanical mode amplitude and polarization conversion effect. This results shows that the main operation limitation in terms of bandwidth arises from the engineering of a proper excitation spectrum of the piezoelectric stack. Changing the drive would allow for operating at any desired conversion frequency at least in the range we tested from 100 kHz to ∼1.4 MHz, giving the advantage of a non-resonant excitation bandwidth with the extremely low power consumption, which in similar devices has produced polarization rotations of roughly 0.07 rad/mV [19].

CONCLUSIONS

In conclusion we have shown a broad band (0–1.4 MHz) polarization conversion effect of near-infrared light using the coherent mechanical motion of an optomechanical metasurface. The all- dielectric, low-loss metasurfaces, potentially integrable at chip level, operates in free-space, and represents a key technology for full control of light parameters (amplitude, phase, and polarization) with the potential for ultra-high frequency operation when high order mechanical modes are excited.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

The device was conceived jointly by SZ, AT, and AP. SZ designed the structure and fabricated the sample starting from a heterostructure grown by GB. MC and AP performed the experiments. All authors participated in the discussion of the results. AP wrote the manuscript with contribution from all the other authors.

FUNDING

Funding from FET-Open project PHENOMEN (GA 713450) was gratefully acknowledged. ICN2 was supported by the Severo Ochoa program from the Spanish MINECO (Grant No. SEV-2017-0706) and funding from the CERCA Programme/Generalitat de Catalunya.

ACKNOWLEDGMENTS

DN-U gratefully acknowledge the support of a Ramón y Cajal postdoctoral fellowship (RYC-2014-15392).

REFERENCES

1. Liu S, Vabishchevich PP, Vaskin A, Reno JL, Keeler GA, Sinclair MB, et al. An all-dielectric metasurface as a broadband optical frequency mixer. *Nat Commun*. (2018) **9**:1–6. doi: 10.1038/s41467-018-04944-9
2. Arbabi A, Horie Y, Bagheri M, Faraon A. Dielectric metasurfaces for complete control of phase and polarization with subwavelength spatial resolution and high transmission. *Nat Nanotech*. (2015) **10**:937. doi: 10.1038/nnano.2015.186
3. Jahani S, Jacob Z. All-dielectric metamaterials. *Nat Nanotechnol*. (2016) **11**:23. doi: 10.1038/nnano.2015.304
4. Jiang J, Fan JA. Global optimization of dielectric metasurfaces using a physics-driven neural network. *Nano Lett*. (2019) **19**:5366. doi: 10.1021/acs.nanolett.9b01857
5. Molesky S, Lin Z, Piggott AY, Jin W, Vucković J, Rodriguez AW. Inverse design in nanophotonics. *Nat Phot*. (2018) **12**:659. doi: 10.1038/s41566-018-0246-9
6. Lee G-Y, Hong J-Y, Hwang SH, Moon S, Kang H, Jeon S, et al. Metasurface eyepiece for augmented reality. *Nat Commun*. (2018) **9**:4562. doi: 10.1038/s41467-018-07011-5
7. Capasso F. The future and promise of flat optics: a personal perspective. *Nanophotonics*. (2018) **7**:953. doi: 10.1515/nanoph-2018-0004
8. Ou J-Y, Plum E, Zhang J, Zheludev NI. An electromechanically reconfigurable plasmonic metamaterial operating in the near-infrared. *Nat Nanotechnol*. (2013) **8**:252. doi: 10.1038/nnano.2013.25
9. Zheludev NI, Plum E. Reconfigurable nanomechanical photonic metamaterials. *Nat Nanotechnol*. (2016) **11**:16. doi: 10.1038/nnano.2015.302

10. Ee H-S, Agarwal R. Tunable metasurface and flat optical zoom lens on a stretchable substrate. *Nano Lett*. (2016) **6:**2818–23. doi: 10.1021/acs.nanolett.6b00618
11. Liu M, Powell DA, Guo R, Shadrivov IV, Kivshar YS. Polarization-induced chirality in metamaterials via optomechanical interaction. *Adv Opt Mater*. (2017) **5**:1600760. doi: 10.1002/adom.201600760
12. Ji R, Hua Y, Chen K, Long K, Fu Y, Zhang X, et al. A switchable metalens based on active tri-layer metasurface. *Plasmonics*. (2019) **14**:165. doi: 10.1007/s11468-018-0789-0
13. Rahmani M, Xu L, Miroshnichenko AE, Komar A, Camacho-Morales R, et al. Reversible thermal tuning of all-dielectric metasurfaces. *Adv Funct Mater*. (2017) **27**:1700580. doi: 10.1002/adfm.201700580
14. Iyer PP, DeCrescent RA, Lewi T, Antonellis N, Schuller JA. Uniform thermos-optic tunability of dielectric metalenses. *Phys Rev Appl*. (2018) **10**:044029. doi: 10.1103/PhysRevApplied.10.044029
15. Yang C, Wang Z, Yuan H, Li K, Zheng X, Mu W, et al. All-dielectric metasurface for highly tunable, narrowband notch filtering. *IEEE Photonics J*. (2019) **11**:4501006. doi: 10.1109/JPHOT.2019.2931702
16. Kamali KZ, Xu L, Ward J, Wang K, Li G, Miroshnichenko AE, et al. Reversible image contrast manipulation with thermally tunable dielectric metasurfaces. *Small*. (2019) **15**:1805142. doi: 10.1002/smll.201805142
17. Bosch M, Shcheerbakov MR, Fan Z, Shvets G. Polarization states synthesizer based on a thermo-optic dielectric metasurface. *J Appl Phys*. (2019) **126**:073102. doi: 10.1063/1.5094158
18. Aspelmeyer M, Kippenberg TJ, Marquardt F. Cavity optomechanics. *Rev Mod Phys*. (2014) **86**:1391. doi: 10.1103/RevModPhys.86.1391
19. Midolo L, Schliesser A, Fiore A. Nano-opto-electro-mechanical systems. *Nat Nanotech*. (2018) **13**:11. doi: 10.1038/s41565-017-0039-1
20. Zanotto S, Tredicucci A, Navarro-Urrios D, Cecchini M, Biasiol G, Mencarelli D, et al. Optomechanics of chiral dielectric metasurfaces. *Adv Opt Mater*. (2019). doi: 10.201810.1002/adom.201901507. [Epub ahead of print].
21. Zanotto S, Mazzamuto G, Riboli F, Biasiol G, La Rocca GC, Tredicucci A, et al. Photonic bands, superchirality, and inverse design of a chiral minimal metasurface. *Nanophotonics*. (2019) **8**:2291. doi: 10.1515/nanoph-2019-0321
22. Adachi S. GaAs, AlAs, and AlxGa1–xAs: material parameters for use in research and device applications. *J Appl Phys*. (1985) **58**:R1. doi: 10.1063/1.336070
23. Collin S. Nanostructure arrays in free-space: optical properties and applications. *Rep Prog Phys*. (2014) **77**:126402. doi: 10.1088/0034-4885/77/12/126402
24. Born M, Wolf E. *Principles of Optics*. 7th ed. Cambridge: Cambridge University Press (1999).
25. Trippe S. Polarization and polarimetry: a review. *JKAS*. (2014) **47**:15. doi: 10.5303/JKAS.2014.47.1.15

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

SOFT MATTER PHYSICS

Dr. Joris Sprakel obtained his PhD under Prof. Martien Cohen Stuart at Wageningen University, for which he received the Polymer Award of the Royal Netherlands Chemical Society. After a postdoctoral fellowship with Prof. David Weitz at Harvard University, he started a research group in the Laboratory of Physical Chemistry and Soft Matter at Wageningen University in 2011, where he is appointed full professor since 2019.His cross-disciplinary team focuses on unravelling the simple rules that underlie molecular complexity in soft and biological materials using an interdisciplinary approach that combines chemistry, instrument development, biology, physics, theory and simulations. In recent years his group has had a particular interest in the question how mechanical cues shape biological processes in plants and plant pathogens, how mechanical patterns in these scenarios can be made visible and how biological approach to constructively respond to mechanical cues can be mimicked in the laboratory using creative mechanochemistry.

Most bonds between molecules become weaker when tension is applied to them. Yet, nature is able to produce bonds that do the opposite: they become stronger under tension. While many examples of such biological catch bonds have been discovered in the last decade, a synthetic realisation has not yet been achieved, mainly because it is not clear what requirements such a synthetic catch bond should meet. In this paper, Sprakel and coworkers present a computational model that shows how the clever combination of simple supramolecular motifs can produce catch bonding behaviour. This model provides guidelines for how a synthetic catch bond can be realized experimentally, starting from common supramolecular motifs.

frontiers in Physics

ORIGINAL RESEARCH
published: 09 September 2020
doi: 10.3389/fphy.2020.00361

Chemical Design Model for Emergent Synthetic Catch Bonds

*Martijn van Galen, Jasper van der Gucht and Joris Sprakel**

Laboratory of Physical Chemistry and Soft Matter, Wageningen University & Research, Wageningen, Netherlands

All primary chemical bonds inherently weaken under increasing tension. Interestingly, nature is able to combine such bonds into protein complexes that accomplish the opposite behavior: they strengthen with increasing tensional force. These complexes known as catch bonds are increasingly considered a general feature in biological systems subjected to mechanical stress. Despite their prevalence in nature however, no truly synthetic realizations of catch bonds have been accomplished so far, as it is a profound challenge to synthetically mimic the allosteric mechanisms employed by protein catch bonds. In this work we propose a computational model that shows how a synthetic catch bond could be accomplished with the help of existing supramolecular motifs and mechanophores, each of which individually act as slip bonds. This model allows us to identify the limits of catch bonding in terms of a number of experimentally measurable parameters. This knowledge could be used to suggest potential molecular candidates, thereby providing a foothold in the ongoing pursuit to realize synthetic catch bonds.

Keywords: catch bond, mechanochemistry, supramolecular chemistry, kinetic Monte Carlo (KMC), chemical kinetics

OPEN ACCESS

Edited by:
Giancarlo Ruocco,
Center for Life NanoScience (IIT), Italy

Reviewed by:
Lining Arnold Ju,
The University of Sydney, Australia
Arlette R. C. Baljon,
San Diego State University,
United States

****Correspondence:***
Joris Sprakel
joris.sprakel@wur.nl

Specialty section:
This article was submitted to Soft Matter Physics, a section of the journal Frontiers in Physics

Received: *30 April 2020*
Accepted: *28 June 2020*
Published: *09 September 2020*

Citation:
van Galen M, van der Gucht J and Sprakel J (2020) Chemical Design Model for Emergent Synthetic Catch Bonds. Front. Phys. 8:361. doi: 10.3389/fphy.2020.00361

1. INTRODUCTION

All primary bonds, covalent or supramolecular, weaken under the action of mechanical load. Such primary chemical bonds, whose dissociation rate grows with increasing force, are known as slip bonds. Interestingly, many protein bonds found in Nature defy this fundamental rule and show the opposite behavior: up to a certain peak force these bonds only strengthen as the applied tensile force grows. Bonds that display this property are known as catch bonds [1]. This behavior is intriguing, as the primary molecular interactions these catch bonding complexes employ each individually act as slip bonds. Catch bonding thus appears to be an emergent phenomenon that occurs when multiple slip bonds work together collectively.

Catch bonds were first discovered experimentally in the cell adhesion protein P-selectin in 2003 [2], and since then a wide range of proteins with mechanical functions have been found to act as catch bonds. Besides the best known examples in biological adhesion proteins such as P-selectin and FimH [3], these include proteins involved in blood clot formation in injured blood vessels, such as the binding between glycoproteins and human von Willebrand Factor [4]. Other examples are membrane proteins involved in mechanotransduction between the extracellular matrix and the intracellular actin skeleton, such as vinculin and integrin [5]. Catch bonding occurs in intracellular applications, such as in myosin-dynein bond formation [6] and in the kinetochore protein machinery that connects microtubules to chromosomes during cell division [7]. More recently, also the membrane bound stator component of flagella was found to display catch bond behavior [8].

Ever since their first discovery, studies have attempted to answer how catch bond proteins manage to achieve this counterintuitive force-induced bond strengthening behavior. Many catch bonds can allosterically switch conformations as a tensile force is applied to them. A prominent example is the mechanism of the FimH catch bond [9, 10]. In this catch bond, the FimH binding domain is closely associated with an autoinhibiting pilin domain that keeps the binding domain in a low-affinity state. When tensile forces pull the two domains apart, the binding domain tightens, which results in a stronger bond [9]. Force-induced conformational switching has also been observed in selectin catch bonds. Selectins contain a ligand-binding lectin domain positioned at an angle relative to an epidermal growth factor domain. Depending on this angle, the selectin catch bond can be in either a weakly bound bent conformation or a more strongly bound extended conformation. Tensile forces applied to selectin catch bonds induce a conformational change from the bent to the extended form, increasing the number of lectin-ligand interactions in the binding pocket [11, 12].

Despite these successes in elucidating the mechanisms of the FimH and selectin catch bonds, directly observing the structural changes in catch bonding proteins remains a challenge in most cases. For this reason, most efforts to characterize catch bond mechanisms have instead focussed on capturing the force response in conceptual physical models [1]. Over the past few years, several models have been applied successfully to a variety of known catch bonds. From a conceptual point of view, the simplest of these energy landscape models is the one state two-pathway model [13]. This model assumes the bond to be in only one state, from which two competing paths of dissociation exist. One of the energy barriers associated with these dissociation paths increases with increasing force, while the other decreases. At low forces, the increasing barrier is rate-limiting, resulting in an increasing bond lifetime as a function of force: catch bond behavior. As the force exceeds a critical value however, the decreasing barrier becomes rate-limiting and the bond reverts back to a slip bond. This catch-to-slip transition is a general feature in biological catch bonds identified so far.

Although the one-state two-pathway model successfully describes catch bond behavior in protein bonds such as those of P-selectin, the fact that it contains an energy barrier that increases with force is a simplification: After all, each of the primary supramolecular interactions employed by the protein individually behave as slip bonds and hence have an energy barrier that decreases with force. This means that the energy barrier in the one-state two-pathway model that increases with force cannot describe a single bond dissociation or reorganization within the protein. Rather, it has to describe a more complex transition, such as the dissociation of a weak supramolecular bond and subsequent formation of a stronger bond. A slightly more complex model that takes into account the fact that the individual bonds weaken with force is the two-state two-pathway model [14, 15]. This model treats the catch bond as a balance between a weakly bound state with a short lifetime and a strongly bound state with a long lifetime (**Figure 1A**). In contrast to the one-state two-pathway model, application of force only lowers each of the energy barriers in the system, but the force affects the height of these barriers to a varying extent. This asymmetry in the force-response "tilts" the energy landscape, which makes occupation of the long lifetime state more likely with respect to the short lifetime state. The net result is an increasing bond lifetime at intermediate forces. As the force increases further, the weakening of both states eventually reduces the bond lifetime again. This model has successfully described catch bonding behavior in biological systems that show transitions between strong and weak states, such as selectin catch bonds [15]. Critically, the two-state two-pathway model provides a physical explanation for the paradox of how a protein complex that can only employ slip bonds as its primary bonds can still behave as a catch bond as a whole.

These insights into the mechanisms of catch bonds beg the question whether it is possible to design synthetic catch bonds that, like FimH and selectin, can switch reversibly between weakly bound and strongly bound states. Such synthetic catch bonds could prove valuable model systems to study the collective effects catch bonds can provide to biological networks and interfaces. An example of such collective effects are the selectin catch bonds found on the surfaces of leukocytes. The catch bond nature of selectin allows these leukocytes to roll on the endothelium surfaces specifically in areas where a fast blood flow exerts shear stresses on the leukocytes, while staying detached when the blood flow is slow [16]. More recent computational studies on nanoparticle networks cross-linked with catch bonds also found enhanced network toughness compared to networks that only utilized slip bonds [17, 18]. The lack of synthetic catch bonds leaves the vast possibility these unique bonds have in shaping material mechanics unexplored for bio-mimetic synthetic materials.

Designing a synthetic realization of a catch bond is no trivial task. The mechanically-induced conformational changes that give rise to catch bonding in proteins are so complex that they cannot be directly mimicked synthetically. However, this does not imply that the concept of a catch bond, a bond that strengthens under tension, is out-of-reach for the synthetic chemist. This paper addresses the question whether a catch bond, separate from its biological reality, can be constructed in a minimal chemical design.

We aim to provide a generic chemical design, as a guide for synthetic chemists, that creates a minimal realization of the conceptual two-state two-pathway model described above. Out of the existing phenomenological catch bond models, this provides the most convenient starting point: it is the only model that explains how a system that consists exclusively of slip bonds can still behave as a catch bond on a collective level, and each of the synthetic building blocks we can employ individually behave as slip bonds.

Here, we propose a minimal design model for a synthetic two-state catch bond, which demonstrates how catch bonding can be achieved by combining mechanophores and supramolecular motifs. Using the conceptual two-state two-pathway model as a foundation, we derive a number of design criteria such a supramolecular construct should fulfill in order to display catch bond activity, and we test these criteria with kinetic monte carlo

simulations. Our results allow us to identify a phase diagram showing a transition between a slip bonding and a catch bonding regime as a function of three control parameters which can be determined experimentally. These findings could be used to suggest potential molecular candidates to make the first true synthetic catch bond a reality, and bring us one step closer to utilizing synthetic catch bonds as model systems to investigate collective effects of catch bonding in biology. Furthermore, such a synthetic catch bond would serve as a novel building block in material science, opening up new avenues to design bio-inspired materials with enhanced properties.

2. CONCEPTUAL DESIGN OF A SYNTHETIC CATCH BOND

To design a synthetic catch bond based on the two-state two-pathway model, we first need to get a conceptual picture of the criteria that must be fulfilled to make such a catch bond work. In the two-state, two-pathways model, a catch bond can transition between two states. The first is a weak, inactivated state I with a short lifetime and the second is a stronger, activated state II with a much longer lifetime. The catch bond relies on the principle that the relative occupancy of these two states changes as a function of force: As the force exerted on the catch bond is increased, the likelihood increases that the bond becomes trapped in the activated state II before it dissociates. As a result, the average bond lifetime of an ensemble of such catch bonds will increase as a function of force.

To obtain a deeper insight in how we can tune the probabilities between states I and II, we can visualize the two-state two-pathway model as an energy landscape (**Figure 1A**) [15]. Two energy minima in this landscape represent the weak, inactivated state I and the strong, activated state II. Bond dissociation is possible from both of these states, and the rates of these processes are governed by energy barriers E_A^1 and E_A^2, respectively. A third energy barrier E_A^{IC} governs the rate of interconversion between states I and II. In the absence of tension, dissociation from the weakly bound state I is more likely than interconversion toward the strong state II, which results in a short bond lifetime. Tension applied to the bond tilts the energy landscape, which lowers the energy barrier for interconversion toward the strong state E_A^{IC} relative to E_A^1. This increases the probability that the system ends up in the strong state II before dissociation, which results in a longer bond lifetime.

This increase in bond lifetime under tension, signaling catch behavior, continues until E_A^{IC} is so low compared to E_A^1 that any further increase in force will not substantially affect the probability to reach state II. If the force is increased beyond this point, the progressive lowering of barriers E_A^1 and E_A^2 only weakens the bond, which makes the system transition from a catch bond to a slip bond at high forces. We note that this catch behavior at low and intermediate forces, and its transition to slip bonding at high forces, is also a feature of all known biological catch bonds, which therefore should be referred to as catch-slip bonds.

Based on this conceptual picture we can reason that the following three criteria must be met to make the catch bond work:

1. E_A^2 should be higher than the energy barrier E_A^1. This criterion ensures that it is more difficult to dissociate from state II than from state I. As a result, the bond lifetime increases as the probability of visiting state II relative to state I increases at greater forces.
2. At low forces, we require that dissociation from the weak state I occurs more quickly than interconversion to the stronger state II. This criterion ensures a short bond lifetime, dictated by the dissociation time of the weak state I. This means that the energy barrier E_A^1 must be substantially lower than the energy barrier E_A^{IC} at low forces.
3. At greater forces however, we require the opposite behavior. Here, dissociation from the weak state I should occur more slowly than interconversion to the stronger state II. This ensures a long lifetime, because the chance of visiting state II before bond dissociation becomes greater. Therefore, energy barrier E_A^1 must be higher than the energy barrier E_A^{IC} at high forces.

The apparent contradiction between the latter two criteria can only be resolved if the height of the barriers E_A^1 and E_A^{IC} scale differently as a function of force. This asymmetry in the response of the kinetic barriers to force is pivotal in the function of the two-state, two-pathway catch bond.

3. MOLECULAR DESIGN MODEL

Now that we have a conceptual picture of the theoretical requirements of a two-state two-pathway catch bond, we set out to develop a molecular design model that fulfills these criteria in order to realize a chemical design for a synthetic catch bond. Our design consists of an oligomeric construct composed of two types of supramolecular interactions (**Figures 1B,C**). The first of these monomers is a "simple" supramolecular slip bond A that dissociates into its unbound state B (**Figure 1B**). The reaction equilibrium of this reaction is governed by the forward rate k_{AB} and reverse rate k_{BA}. In the inactive, weakly bound state I of the catch bond, only these simple slip bonds A are engaged.

Our two-state, two-pathway catch bond also requires the ability to switch to a more strongly bound state II under force in order to fulfill criteria 1, 2, and 3 we have identified above. To this end, we include a second type of force-responsive monomer in our supramolecular oligomer. This monomer is a so-called mechanophore: a molecule that consists in an inactive state C at rest and undergoes a force-induced conformational change k_{CD} that allows it to form a supramolecular slip bond D (**Figure 1C**). The formation of this supramolecular slip bond D corresponds to the activated state II of the catch bond. In turn, this slip bond D reversibly dissociates into the unbound state E with forward and backwards rates k_{DE} and k_{ED} respectively. As the formation of the supramolecular bond stabilizes the active state of the mechanophore, deactivation to state C occurs exclusively via the debonded state E at reaction rate k_{EC}.

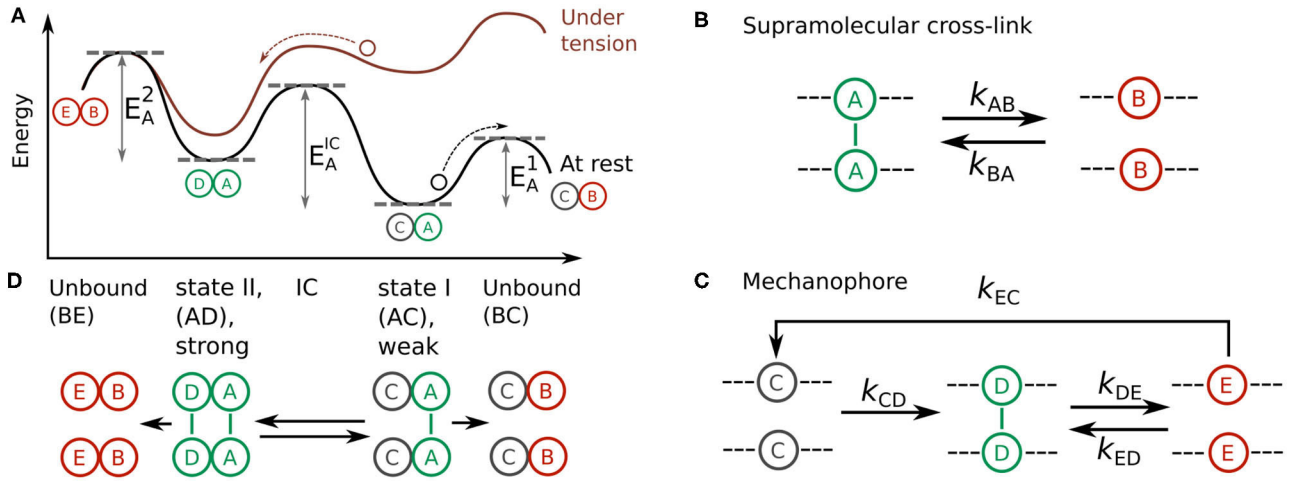


FIGURE 1 | (A) Energy landscape of a two-state, two-pathway catch bond at rest (black) and under tension (red). The activation energies of dissociation from state I, dissociation from state II, and interconversion between these states are shown as E_A^1, E_A^2, and E_A^{IC}, respectively. **(B,C)** Design model of a synthetic catch bond that consists of **(B)** a supramolecular cross-link and **(C)** a mechanophore capable of forming a supramolecular cross-link under force-activation. **(D)** Schematic representation of a catch bond consisting of one supramolecular cross-link and one mechanophore in weak state I (AC) and strong state II (AD), along with the two unbound states BC and BE that could be reached from these states.

The catch-bonding oligomer is constructed of $N = N_A + N_C$ monomers. Let the symbols n_i with $i = A, B..E$ denote the number of monomers in their respective conformational state, with the condition $\Sigma_i n_i = N$. Since the states A and D form supramolecular bonds, only these contribute to the mechanical stability of the construct, so that the total number of bonds equals $n_{bonds} = n_A + n_D$. The overall picture of this catch bond design is as follows (**Figure 1D**): in the absence of applied force, the oligomer exists in the inactivated state I with all N_A monomers in the bound (A) state and all mechanophores N_C in the inactivated state C. In this inactivated state of the catch bond, the bond lifetime is thus governed by the weak slip bonds A. Application of small stretching forces increases the probability of activating the $C \rightarrow D$ bonds relative to the probability of rupturing some $A \rightarrow B$ bonds. This pushes the system into the activated state II where the bond lifetime is governed by the D supramolecular interactions. Provided that the bond strength of the D bonds is greater than that of the A bonds, we can expect state II to be a stronger overall bond than state I. In short, this means that we can fulfill the criterion 1 of a two-state two-pathway catch bond ($E_A^2 > E_A^1$) by ensuring that the bond strength of the D bonds is greater than the bond strength of the A bonds.

We note that the activated state II is a multivalent state, as it is comprised of both $A - A$ and $D - D$ interactions. As a result, dissociation from state II is possible via multiple pathways, in which these interactions break either simultaneously or sequentially. This makes the overall energy landscape of the proposed catch bond slightly more complex as compared to a monovalent two-state catch bond. In spite of this more complicated picture, the system can achieve catch bonding as long as the lifetime of the multivalent state II is longer than the life time of state I.

To ensure that this design model also fulfills the second and third criteria of a two-state two-pathway catch bond, we have to consider the kinetics of the system. In our catch bond design, we can write out the following set of kinetic equations that govern the occupation of the states over time:

$$\frac{dn_A}{dt} = -k_{AB} \cdot n_A + k_{BA}\,(N_A - n_A) \quad (1)$$

$$\frac{dn_C}{dt} = -k_{CD} \cdot n_C + k_{EC}\,(N_C - n_C - n_D) \quad (2)$$

$$\frac{dn_D}{dt} = k_{CD} \cdot n_C - k_{DE} \cdot n_D + k_{ED}\,(N_C - n_C - n_D) \quad (3)$$

As discussed above, criteria 2 and 3 of a two-state, two-pathway model dictate that the height of the energy barrier of dissociation from the inactive state I should be lower than the height of the barrier of interconversion between the inactive state I and the active state II at small forces, but higher at large forces. Given that in our case the inactive state I is a state with predominantly A bonds (high n_A) and the active state II is the state with predominantly D bonds (high n_D), this can be achieved if the rate constants k_{AB} and k_{CD} are force-dependent and scale differently as a function of force. We incorporate the effect of applied force in our system by presuming that the rate constants of supramolecular bond rupture ($A \rightarrow B$ and $D \rightarrow E$) and mechanophore-activation ($C \rightarrow D$) display Kramers-Bell type thermally-activated and mechanically-enhanced kinetics [19]. For these reaction steps, the rate constants can then be written as:

$$k_{ij} = k_{ij,0} \cdot \exp[\beta f \delta x_i] \quad (4)$$

Here, $(ij) = (AB)$, (CD), and (DE), $\beta = 1/k_B T$, $k_{ij,0}$ is the reaction rate at zero force and δx_i is the activation length. This activation length δx_i denotes the length change of the bond in

the activated state relative to the bond at rest, and determines the susceptibility of the bond to mechanical activation. The rate constants of all other transitions are assumed to be independent of the applied force. According to the Kramers-Bell model, energy barriers are exponentially lowered as a function of force, and the degree to which they do so depends on the activation length δx_i. If supramolecular bond A and mechanophore C are chosen such that $k_{AB,0} > k_{CD,0}$ and $\delta x_A < \delta x_C$, we ensure that $k_{AB} > k_{CD}$ at small forces, while $k_{AB} < k_{CD}$ at large forces. Hence, design criteria 2 and 3 are both met. Within this argument based on rate constants, design criterion 1 ($E_A^2 > E_A^1$) can be expressed as $k_{AB} > k_{DE}$. This can be realized if $k_{AB,0} > k_{DE,0}$, assuming that δx_A is similar to $\delta\, x_D$.

In essence, the synthetic catch bond we propose exists as a balance between two subpopulations: a weakly bound state I from which dissociation is slow and a more strongly bound state II from which dissociation is faster. Meeting criteria 2 and 3 ensures that increasing the force shifts the balance between these two subpopulations toward the more strongly bound state II. This force-induced shifting between subpopulations is the key mechanism by which many relevant biological two-state catch bonds have been identified to work, such as the interaction between von Willebrand factor and platelet glycoprotein Ibα [20], and P-selectin [15].

We note in passing that the Kramers-Bell description of force-enhanced thermal barrier transitions does not take into account the directionality of the applied force. Rather, in the catch bond we propose (**Figure 1D**) it is assumed that a fixed magnitude of force is applied to cross-links ($A-A$ and $D-D$) and the mechanophore (C). In practice however, the dissociation of the $A-A$ and $D-D$ bonds might require a different force directionality than the activation of the mechanophore $C \rightarrow D$. For example, activation of mechanophore C could require the rupture of intramolecular bonds within C oriented in a different direction than the intermolecular cross-link $A-A$. Depending on the direction in which the force is applied to the catch bond as a whole, the magnitude of the applied tensional load might be distributed differently in each of the relevant directions, which could cause cross-link C to experience a different magnitude of force than supramolecular bonds $A-A$ and $D-D$. We do not account for this effect in our simplified design model. Moreover, the Kramers-Bell description (Equation 4) assumes that the applied force only affects the height of the energy barrier for bond dissociation, but does not affect its curvature [21]. This assumption is only valid if the barrier is sufficiently steep, that is, if the activation length is comparatively short with respect to the barrier height, which is the case for most mechanically stable supramolecular motifs.

4. RESULTS

To test whether our molecular design model displays catch bonding, we solve the system of kinetic equations (1–3) with a Gillespie Kinetic Monte Carlo (KMC) algorithm, which takes into account the effect of thermal fluctuations, as discussed in [22]. We run a simulation until $n_{bonds} = n_A + n_D = 0$, which signifies rupture of the oligomer. For each of the designs and loading condition we test, we run 1,000 independent simulations to ensure sufficient statistics. We simulate these rupture events in the limit of constant strain. Since the supramolecular bonds (A-A and D-D) that connect two unimers act as springs, the condition of constant strain implies a constant force f on each of the supramolecular bonds in parallel. As a result, the force per bond f is time-invariant in the limit of constant strain, while the cumulative load on the dimer $F(t) = fn_{bonds}$ varies over time.

We first consider the simplest case of our supramolecular oligomer, which is a dimer composed of only a supramolecular slip bond A-A without mechanophores: $N = N_A$ and $N_C = 0$. For a single supramolecular bond ($N_A = 1$), we observe Poisson statistics, where the dimensionless bond lifetime $\tilde{\tau} = \tau k_{BA,0}$ decays exponentially as a function of the dimensionless force $\tilde{f} = f\delta x_A/k_BT$ (**Figure 2A**). Each data point in **Figure 2A** is averaged over 10,000 simulations per force. For multivalent dimers ($N_A > 1$), we observe two exponential decay regimes in the bond lifetime. This effect was previously explained for failure of multivalent colloidal aggregates [23]. For small forces, rupture of the multivalent dimer is cooperative. Here, dissociation occurs relatively slowly compared to reassociation ($k_{AB} << k_{BA}$). As a result, dissociated bonds within the dimer have ample time to reform as long as other bonds within the dimer are still present, preventing dissociation of the dimer as a whole. Due to this multivalency effect, one expects an exponential slope that scales with N_A. Indeed, we find that increasing N_A strongly increases the slope of the bond lifetime at small forces. By contrast, at large forces we observe a slope that is independent of N_A. Here, dissociation occurs substantially faster than association ($k_{AB} >> k_{BA}$), so that bond rupture is simply a sequence of rupturing all bonds in succession. In this regime, adding more bonds N_A would have little effect on the bond lifetime, so that one would indeed expect the slope to be insensitive to N_A. Due to this force-dependent transition from cooperative to non-cooperative behavior, the degree to which the rupture rate is affected by the force decreases as a function of force. This multivalency effect is a useful tool in designing a synthetic catch bond, because it could help the rate of mechanophore activation (k_{CD}) overtake the dissociation of the A-A bonds (k_{AB}) more easily at large forces.

We now introduce a mechanophore into our dimer to explore the possibility to create a catch bond. The simplest catch bond we can consider consists of one A-A dimer and one mechanophore ($N_A = N_C = 1$). We choose our kinetic parameters such that they fulfill the three criteria of a two-state two-pathway catch bond: $k_{DE,0} < k_{AB,0}$ (criterion 1), $k_{AB,0} > k_{CD,0}$ (criterion 2), and $dx_A < dx_C$ (criterion 3). The bond reformation rates k_{BA} and k_{ED} are controlled by diffusion and thus considered of similar magnitude,so that $k_{BA} = k_{ED} = 1$. As a simplification we assume that the spontaneous conversion of the activated to the inactivated chromophore is slow compared to the other rates in the system, so that it can be ignored ($k_{EC} = 0$).

Interestingly, we find that adding a single mechanophore to one A-A dimer bond already yields convincing catch bonding behavior (**Figure 2B**, averaged over 1,000 simulations per force). We can understand how this occurs by studying the average

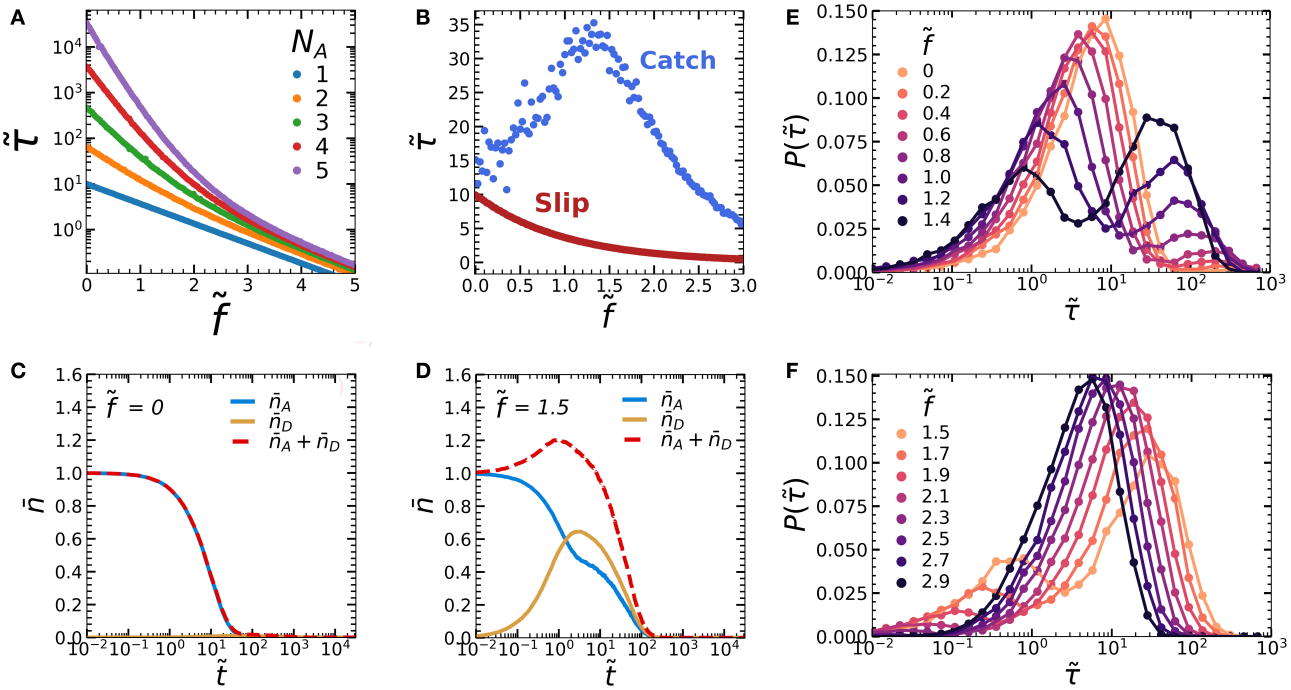


FIGURE 2 | (A) Force-lifetime curve of an oligomer consisting of a varying number of subunits (N_A), without a mechanophore ($N_C = 0$). **(B)** Force lifetime curves of a single $N_A = 1$ oligomer displays slip bond behavior in the absence of a mechanophore (red curve, $N_C = 0$) and catch bond behavior in the presence of a mechanophore (blue curve, $N_C = 1$). **(C,D)** Kinetic diagrams of the blue curve in figure b showing the average number of bonds in the A and D states ($\bar{n}_A$ and $\bar{n}_D$) over simulation time steps $\tilde{t}$ at $\tilde{f} = 0$ **(C)** and around maximum lifetime, at $\tilde{f} = 1.5$ **(D)**. **(E,F)** Logarithmically binned histograms of the bond dissociation times $\tilde{\tau}$ at different forces for **(E)** $\tilde{f} < 1.5$ and **(F)** $\tilde{f} > 1.5$. Simulations were run with zero-force rate constants $k_{AB,0} = 0.1$, $k_{CD,0} = 0.001$, $k_{DE,0} = 0.01$ and activation lengths $\delta x_C = 4.6$, and $\delta x_D = \delta x_A = 1$.

number of bonds in the A-A and D-D states over time (**Figures 2C,D**). At $\tilde{f} = 0$, we find that almost exclusively the weak A-A bonds are present, while only a small number of the mechanophores activate due to thermal activation (**Figure 2C**). By contrast, under tension at the maximum of the force-lifetime curve $\tilde{f} = 1.5$, the strong mechanical scaling of the mechanophore activation rate increases k_{CD} relative to the A-A dimer rupture rate k_{AB}. This causes an increasing amount of oligomers to become trapped in the stronger, activated D-D state before rupture (**Figure 2D**). In the latter case we also find that the A-A dissociation ($\bar{n}_A$) occurs in two stages: After a quick initial decay, the dissociation rate slows down dramatically as soon as $\bar{n}_D$ increases. We can attribute this to a multivalency effect: as soon as both the A-A and D-D dimers are formed, the A-A bond can dissociate without breaking the dimer. This allows the bond to reform via rate constant k_{BA}. Due to this reassociation, we can expect $\bar{n}_A$ to drop more slowly.

All data we have discussed so far are averaged over at least 1,000 simulations per datapoint. While such averaged properties can inform us whether a population of bonds display catch bonding as a whole, they provide no information on the lifetime distribution at the level of the individual bonds. To obtain such insight, we constructed logarithmically binned histograms of the bond dissociation time $\tilde{\tau}$ for varying forces in the regime of increasing bond lifetime, $\tilde{f} < 1.5$ (**Figure 2E**) and in the regime of decreasing bond lifetime $\tilde{f} > 1.5$ (**Figure 2F**). At force $\tilde{f} = 0$, bond dissociation occurs from a single population. As we increase the force toward $\tilde{f} = 1.5$, we observe that an increasing fraction of bonds dissociates from a second population with a higher lifetime. Simultaneously, the lifetimes of both these populations decrease with increasing force. This trend continues as we increase the force beyond $\tilde{f} > 1.5$ as shown in **Figure 2F**, until only the long-lifetime population is left at $\tilde{f} = 2.3$. In short, the lifetime enhancement as seen in the catch bond force-lifetime curve in **Figure 2B** is only visible as a collective effect. On the level of the individual bonds however, dissociation proceeds from two populations, and the lifetimes of each of these populations are in fact lowered by the applied force. The increasing average bond lifetime is thus caused by a shift in probability between two states, rather than by a increase in the lifetimes of the states themselves. These observations are in line with the two-state, two pathway model.

We have now established that catch bonding can be achieved in our model system by combining a simple oligomer and a mechanophore. However, the phase space for synthetic design is vast, and the tunability of the kinetic parameters depends on the availability of supramolecular motifs and mechanophores. For a more complete picture, we thus set out on a systematic exploration of the parameter space. Although we have previously reasoned that a certain range in the parameters $k_{AB,0}/k_{CD,0}$,

$\delta x_C/\delta x_A$, and $k_{AB,0}/k_{DE,0}$ is required to achieve catch bonding, the question remains where the boundaries of this catch bonding regime lie exactly. Furthermore, we can wonder how different aspects of our catch bond can be tuned by these parameters, such as the relative increase in bond lifetime under force or the location of the optimum of our force-bond lifetime curve. Such knowledge would be especially useful in the rational design of artificial catch bonds.

To answer these questions, we systematically vary each of the parameters $k_{AB,0}/k_{CD,0}$, $\delta x_C/\delta x_A$, and $k_{AB,0}/k_{DE,0}$ while recording their force-lifetime curves (**Figure 3**). First, we vary the ratio of activation lengths between the process of mechanophore activation and A-A bond rupture $\delta x_C/\delta x_A$, by varying δx_C at constant $\delta x_A = 1$ (**Figure 3A**). When the activation lengths are similar ($\delta x_C/\delta x_A = 1$), we find clear slip-bond behavior. As we increase $\delta x_C/\delta x_A$, we find a transition from slip bonding to catch bonding behavior, paired with an increasing lifetime at intermediate forces. At greater values of $\delta x_C/\delta x_A$, k_{CD} becomes greater than k_{AB} at intermediate forces, which increases the chance the mechanophore activates and forms the D-D bond before rupture of the A-A slip bond occurs. At the boundary between the slip and catch regimes, we observe a case of ideal bonding at ($\delta x_C/\delta x_A = 3.1$) where the lifetime $\tilde{\tau}$ remains constant for a sizeable range of forces $0 < \tilde{f} < 1$. In short, we can increase $\delta x_C/\delta x_A$ to increase the fraction of bonds that reach the activated state II before dissociation from state I. This allows us to tune the average force-lifetime curve between slip bonding, ideal bonding and catch bonding.

When we vary $k_{AB,0}/k_{CD,0}$ by changing $k_{CD,0}$, we observe an altogether different effect on the catch bond curve (**Figure 3B**). In the limit $k_{AB,0}/k_{CD,0} \to 0$ the kinetic equations dictate that the catch bond will be universally activated even at $\tilde{f} = 0$, which results in slip bond behavior governed by the bond lifetime of the strong D-D bond (k_{DE}). We indeed observe a force-lifetime curve that tends toward a slip bond at low values of $k_{AB,0}/k_{CD,0}$. As $k_{AB,0}/k_{CD,0}$ increases, the bond lifetimes at low forces and at the force maximum drop consistently, while the force maximum itself shifts toward greater forces. This tunability of the force maximum is an interesting feature which could be exploited in the design of artificial catch bonds. It can be explained as follows: At high values of $k_{AB,0}/k_{CD,0}$, the initial offset in rate constants k_{CD} and k_{AB} is greater, which means that greater forces are required before the rate of mechanophore activation overtakes the rate of A-A bond rupture, activating the catch bond. Finally, the bond lifetime at large forces appears unaffected by $k_{AB,0}/k_{CD,0}$. At such large forces the catch bond will be universally activated, which means that further changes in $k_{AB,0}/k_{CD,0}$ will have little effect. In short, $k_{AB,0}/k_{CD,0}$ can be used to strongly tune the lifetime at small forces and the force at which the maximum lifetime is reached, without affecting the lifetime at large forces. The effect of $k_{AB,0}/k_{CD,0}$ can also be understood in terms of our conceptual model: Increasing $k_{AB,0}/k_{CD,0}$ lowers the energy barrier E_A^1 relative to barrier E_A^{IC} at low forces. This lowers the bond lifetime, because an increasing fraction of bonds dissociate from state I without reaching activated state II. At large forces beyond the lifetime maximum however, the chance of reaching state II is high regardless of $k_{AB,0}/k_{CD,0}$, since $\delta x_C > \delta x_A$. When most catch bonds reach state II, the bond lifetime is governed by barrier E_A^2. This barrier is unaffected by $k_{AB,0}/k_{CD,0}$ and hence the lifetime at large forces is unaffected by $k_{AB,0}/k_{CD,0}$.

However, varying the ratio of the initial bond A-A and D-D bond strengths $k_{AB,0}/k_{DE,0}$ (**Figure 3C**) has a large effect on the bond lifetime at large forces. The lifetime at the force maximum is also strongly affected, while the lifetime at low forces is unaffected. As we increase the value of $k_{AB,0}/k_{DE,0}$, we find a clear transition from slip bond behavior to catch bond behavior with increasing bond lifetimes. At low $k_{AB,0}/k_{DE,0}$, the mechanophore might be activated, but this activation does not lead to the formation of a stronger bond. As such, the lifetime of the construct as a whole remains governed by the weak A-A cross-link, and we observe slip bond behavior. For larger $k_{AB,0}/k_{DE,0}$ on the other hand, activation of the mechanophore leads to an increasingly strong D-D bond and hence a longer bond lifetime at the force maximum and larger forces. An interesting observation is that some catch bonding can already be observed before the D-D bond is stronger than the A-A bond, at $k_{AB,0}/k_{DE,0} = 1$. If the A-A and D-D bond are of comparable strength, we can expect the lifetime of the activated catch bond as a whole to increase due to multivalency. This has interesting implications for the effort to create an artificial catch bond, as it means that it is not necessary for the mechanophore to form a very strong D-D cross-link. Rather, it is sufficient if the activated state contains multiple bonds of a similar strength as the bonds in the inactivated state. In terms of the conceptual two-state, two-pathway picture, increasing $k_{AB,0}/k_{DE,0}$ increases the energy barrier E_A^2 relative to barrier E_A^1. This increases the bond lifetime in the activated state II relative to state I but has little effect on the chance of reaching state II. As a result, we find an increasing bond lifetime at high forces, where the chance of reaching state II is high, but little increase in lifetime at low forces, where the chance of reaching state II is low.

Combined, these results show how different aspects of our catch bond can be tuned by varying three parameters. Specifically, we can tune the location of the force maximum and the bond lifetime at low force by varying $k_{AB,0}/k_{CD,0}$; we can tune the bond lifetime at large forces by varying $k_{AB,0}/k_{DE,0}$, and we can tune the lifetime of our catch bond at the force maximum by varying all three parameters. To also obtain quantitative insight in the minimal conditions required to achieve catch bonding, we can define a parameter that quantifies the degree of catch bonding and study how this parameter varies as a function of our three control parameters. We define the parameter $\Delta\tilde{\tau}/\tilde{\tau}$, which denotes the mechanical enhancement: the increase of the bond lifetime at the force maximum relative to the bond lifetime in the absence of force (**Figure 4A**, inset). $\Delta\tilde{\tau}/\tilde{\tau}$ is equal to 0 for slip bonds and ideal bonds, and increases with more pronounced catch-bonding.

We calculated $\Delta\tilde{\tau}/\tilde{\tau}$ for each of the three simulation series we discussed above, as shown in **Figures 4A–C**. We observe that increasing $\delta x_C/\delta x_A$ leads to a linear increase in the mechanical enhancement above a critical value $\delta x_C/\delta x_A \approx 2$. Catch bonding increases logarithmically as $k_{AB,0}/k_{CD,0}$ and $k_{AB,0}/k_{DE,0}$ both

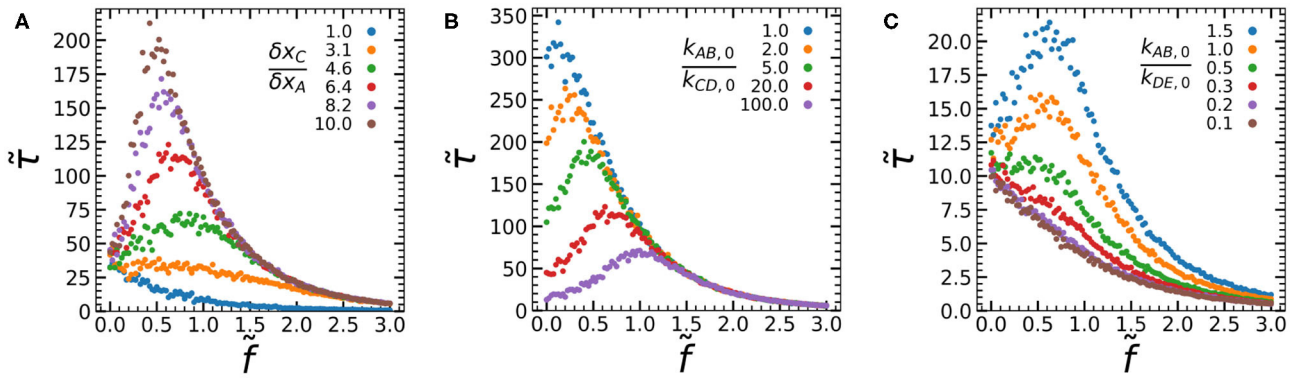


FIGURE 3 | Force-lifetime curves simulated at varying $\delta x_C/\delta x_A$ **(A)**, $k_{AB,0}/k_{CD,0}$ **(B)**, and $k_{AB,0}/k_{DE,0}$ **(C)**. Data was recorded at $\delta x_C/\delta x_A = 6.4$ **(B,C)**, $k_{AB,0}/k_{CD,0} = 20$ **(A,C)**, and $k_{AB,0}/k_{DE,0} = 10$ **(A,B)**. All data are averages of 1,000 simulations per force, at $N_A = N_C = 1$, $dx_{DE} = dx_{AB} = 1$, $k_{ED,0} = k_{BA,0} = 1$, and $k_{AB,0} = 0.1$.

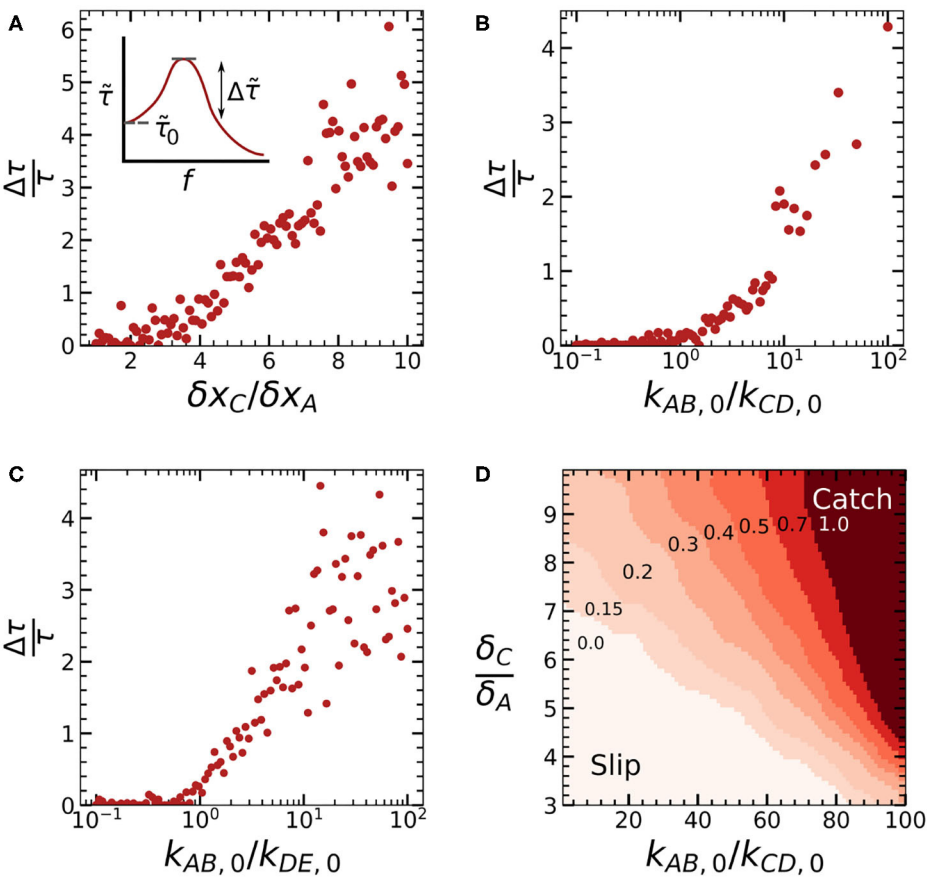


FIGURE 4 | Plots of the mechanical enhancement $\Delta\tilde{\tau}/\tau$ [**(A)**, inset] as a function of the three control parameters $\delta x_C/\delta x_A$ **(A)**, $k_{AB,0}/k_{CD,0}$ **(B)**, and $k_{AB,0}/k_{DE,0}$ **(C)** calculated for the series of force-lifetime curves shown in **Figure 3**. **(D)** Contour plot of $\Delta\tilde{\tau}/\tau^0$ as a function of both $\delta x_C/\delta x_A$ and $k_{AB,0}/k_{CD,0}$.

increase beyond a critical value of $\approx$ 1, although some catch bonding can already be observed for $k_{AB,0}/k_{DE,0} < 1$, due to the multivalency effect we discussed above. The parameters $k_{AB,0}/k_{CD,0}$ and $\delta x_C/\delta x_A$ together determine whether criteria 2 and 3 of a two-state, two-pathways model are met. This means that we can expect the greatest effect on the mechanical enhancement if both parameters are increased together. To study the combined effect of these two parameters, we carried out a range of simulations where we varied both $k_{AB,0}/k_{CD,0}$ and $k_{AB,0}/k_{CD,0}$. As a result, we obtained a gaussian-binned phase

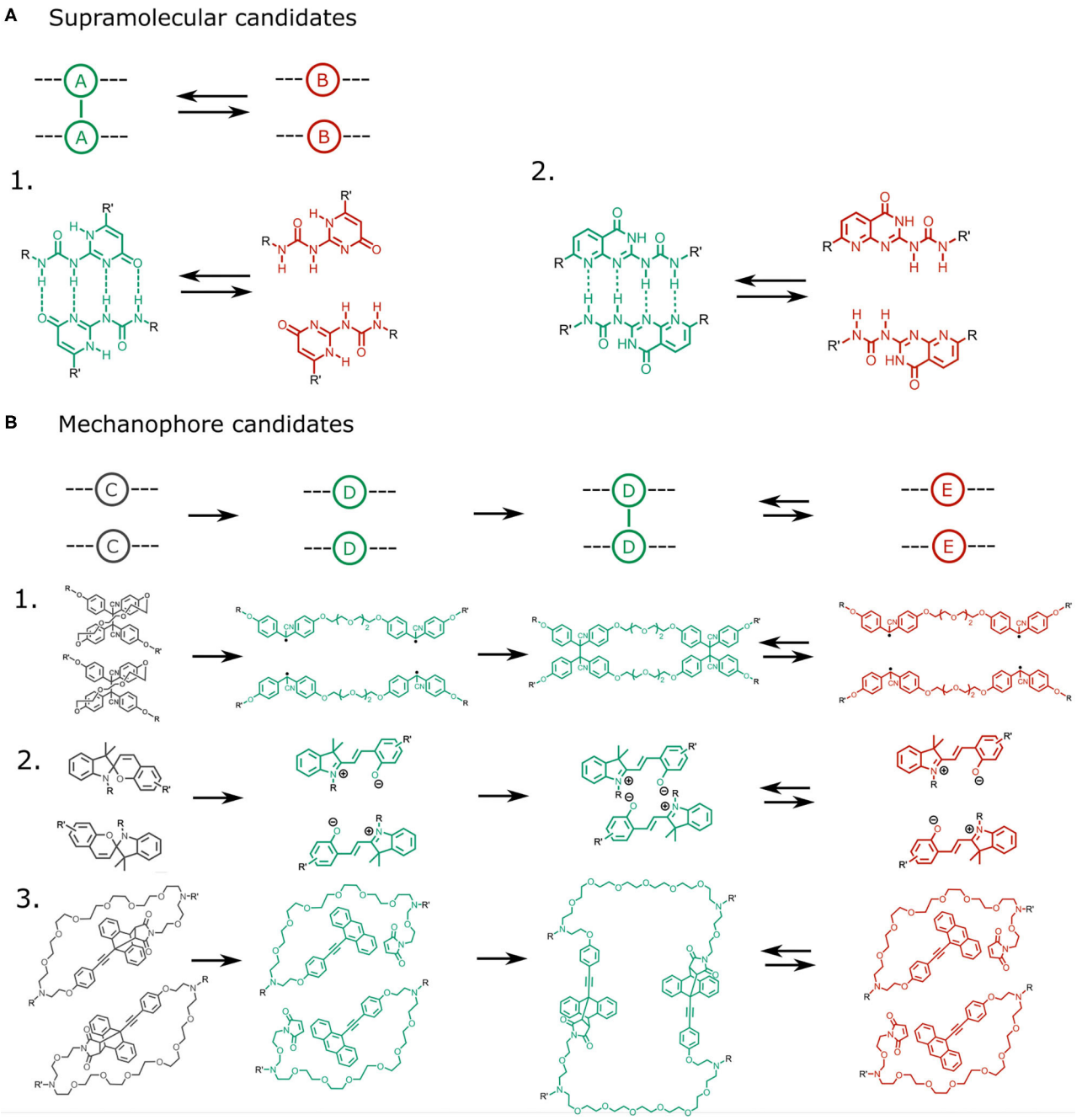


FIGURE 5 | (A) Potential quadruple hydrogen bonding candidates for the supramolecular interaction $A - A$. Two examples based on 2-ureido-4[1H]-pyrimidone are shown [24, 25]. **(B)** Potential mechanophore candidates should be capable of converting an intramolecular bond (C) into an intermolecular bond ($D - D$). Hypothetical examples are: 1. a bridged tetraphenyl succinonitrile molecule containing a reversibly cleavable C-C bond [26, 27], 2. spiropyran, which converts into zwitterionic merocyanine after rupture of an intramolecular C-C bond, allowing it to dimerize through electrostatic interactions and $\pi - \pi$ stacking [28–31], and 3. Diels-Alder adducts, which can dissociate through a mechanically induced retro-Diels-Alder reaction, and subsequently reassociate to form intermolecular Diels-Alder bonds [32].

diagram (**Figure 4D**) showing a gradual transition from a clear slip-regime at low $k_{AB,0}/k_{CD,0}$ and low $\delta x_C/\delta x_A$ to a regime which shows distinct catch bond behavior at high $k_{AB,0}/k_{CD,0}$ and $\delta x_C/\delta x_A$.

5. DISCUSSION

In this work, we have presented a minimal chemical design model for a synthetic catch bond. This model provides a

theoretical picture of how synthetic catch bonding could be achieved in an oligomer consisting of supramolecular moieties (A) and mechanophores (C), each of which individually act as slip bonds. We can tune the characteristic force-lifetime curve of these catch bonds with three kinetic control parameters derived from the energy landscape of the two-state two-pathway model. Specifically, we can tune the lifetime at zero force with $k_{AB,0}/k_{CD,0}$, the lifetime in the activated state with $k_{AB,0}/k_{DE,0}$, and the lifetime at the force maximum with the ratio of the activation lengths $\delta x_C/\delta x_A$ as well as the previous two parameters. These parameters might therefore offer a foothold to not only build a synthetic catch bond, but also tune its characteristics.

We have determined the limits of the range these parameters should take in order to achieve catch bond behavior by quantifying the bond lifetime enhancement at the force maximum relative to the bond lifetime at zero force. These findings have allowed us to identify a phase diagram of catch bond behavior that could form a theoretical framework. With this framework in hand, the next step is to predict potential molecular candidates to fulfill the role of our supramolecular cross-link (A) and mechanophore (C). Theoretically, many primary supramolecular bonding types can be considered for the supramolecular bond A, such as those based on hydrogen bonds, hydrophobic interactions or electrostatic interactions, for which a wide variety of molecular realizations are available to the supramolecular chemist (**Figure 5A**).

The choice for the mechanophore C requires more attention, as the suitable molecule should be capable of a force-induced transition from an intramolecular to an intermolecular bond, for which several candidates are available (**Figure 5B**). One potential realization of molecule C is spiropyran, which can undergo a reversible covalent bond dissociation [29] converting it into the zwitterionic merocyanin molecule capable of dimerization through $\pi - \pi$ stacking and electrostatic interactions [30, 31]. Other candidates include Diels-Alder adducts of aromatic molecules, which can be activated by a mechanically-triggered retro Diels-Alder reaction [32]. In a Diels-Alder mechanophore, force-induced dissociation of an intramolecular Diels-Alder adduct can trigger the formation of an intermolecular Diels-Alder adduct or the exposure of an aromatic group capable of undergoing pi-bonding. Finally, succinonitrile compounds may be considered. These contain labile covalent bonds that can dissociate reversibly into radicals, whose recombination with adjacent moieties can trigger intermolecular bonding from a labile intramolecular bond [26, 27].

In selecting a suitable supramolecular unit/mechanophore pair, care should be taken that these satisfy the criteria we have identified. The criterion that $k_{AB,0}/k_{CD,0} >> 1$ should be viable as many mechanophore activations rely on the dissociation of covalent bonds, whereas the supramolecular cross links rely on weaker interactions. Similarly, the criterion that $k_{AB,0}/k_{DE,0} >> 1$ should be viable as the bonds formed upon mechanophore activation are generally stronger than the supramolecular interactions, especially in the case of diels alder adducts and succinonitrile compounds. However, one of the greatest bottlenecks in selecting a suitable A and C pair is likely to ensure a sufficiently large activation length ratio $\delta x_C/\delta x_A$. Our simulations reveal that $\delta x_C/\delta x_A$ must be greater than 3 to obtain a catch bond. However, experimental evidence reveals that activation lengths of many mechanophores are in fact similar to those of many supramolecular interactions. For example, the activation length of the spiropyran ring opening is around 1.93 Å [28], which does not exceed the activation length of the commonly used supramolecular unit 2-ureido-4[1H]-pyrimidone (2.3 Å) [24].

A potential solution to this challenge is to make use of the force geometry. While our model does not take into account the effect of molecular groups surrounding the mechanophore, experimental results show the effective activation length of mechanophores can depend strongly on the surrounding chemical environment. For example, single molecule force spectroscopy studies on a mechanophore embedded in a polymer show that the activation length of a mechanophore can be enhanced by the surrounding polymer backbone [33]. In this case, the activation length is interpreted as the effective contour length difference between the mechanophore at rest and in its activated state in the direction of the applied force. This suggests that by making clever use of the way in which force is distributed over the bond, it is perhaps possible to achieve an effective activation length difference without employing exotic molecules with intrinsically large activation lengths. This might also offer an explanation on how natural proteins can act as catch bonds in spite of the fact that they mostly employ common supramolecular interactions.

Another avenue that could be explored in the future is the effect of multivalency in achieving catch bond behavior. In this work, we have only looked at multivalency in the number of supramolecular (A) units, but multivalency in the number of mechanophores (C) might also have important implications: In the limit of constant strain, we could predict that incorporating multiple mechanophore units increases the collective work applied to the bond as $n_C f \delta x$. Under constant tension force, this means that the effect of the activation length of the interconversion between the inactive and active states is multiplied by n_C. We can imagine this as follows: as soon as one of the mechanophores activates, the lifetime of the bond is increased, which in turn increases the likelihood that other mechanophores are activated before the catch bond dissociates as a whole. This effect might allow catch bonds to be made out of mechanophores that otherwise would not have a sufficiently large activation length. It could also have implications for biological catch bonds. Despite having only a limited pool of amino acids and thus interactions to choose from, catch bonding protein complexes could employ multivalency to obtain large effective activation lengths for the transitions between their inactive and active states.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

JS conceived the project. JG and MG wrote the simulation code. MG performed the simulations and the data analysis. All authors discussed the data and their interpretation and co-wrote the manuscript.

FUNDING

The research presented in this article was financially supported by VLAG Graduate School. JG acknowledges the European Research Council for financial support (ERC CoG SOFTBREAK).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2020.00361/full#supplementary-material

REFERENCES

1. Thomas WE, Vogel V, Sokurenko E. Biophysics of catch bonds. *Annu Rev Biophys.* (2008) **37**:399–416. doi: 10.1146/annurev.biophys.37.032807.125804
2. Marshall B, Long M, Piper J, Yago T, McEver R, Zhu C. Direct observation of catch bonds involving cell-adhesion molecules. *Nature.* (2003) **423**:190–3. doi: 10.1038/nature01605
3. Thomas W, Nilsson L, Forero M, Sokurenko E, Vogel V. Shear-dependent 'stick-and-roll' adhesion of type 1 fimbriated *Escherichia coli*. *Mol Microbiol.* (2004) **53**:1545–57. doi: 10.1111/j.1365-2958.2004.04226.x
4. Yago T, Lou J, Wu T, Yang J, Miner JJ, Coburn L, et al. Platelet glycoprotein lb alpha forms catch bonds with human WT vWF but not with type 2B von Willebrand disease vWF. *J Clin Investig.* (2008) **118**:3195–207. doi: 10.1172/JCI35754
5. Huang DL, Bax NA, Buckley CD, Weis WI, Dunn AR, Vinculin forms a directionally asymmetric catch bond with f-actin. *Science.* (2017) **357**:703–6. doi: 10.1126/science.aan2556
6. Rai AK, Rai A, Ramaiya AJ, Jha R, Mallik R. Molecular adaptations allow dynein to generate large collective forces inside cells. *Cell.* (2013) **152**:172–82. doi: 10.1016/j.cell.2012.11.044
7. Akiyoshi B, Sarangapani KK, Powers AF, Nelson CR, Reichow SL, Arellano-Santoyo H, et al. Tension directly stabilizes reconstituted kinetochore-microtubule attachments. *Nature.* (2010) **468**:576–U255. doi: 10.1038/nature09594
8. Nord AL, Gachon E, Perez-Carrasco R, Nirody JA, Barducci A, Berry RM, et al. Catch bond drives stator mechanosensitivity in the bacterial flagellar motor. *Proc Natl Acad Sci USA.* (2017) **114**:12952–7. doi: 10.1073/pnas.1716002114
9. Trong IL, Aprikian P, Kidd BA, Forero-Shelton M, Tchesnokova V, Rajagopal P, et al. Structural basis for mechanical force regulation of the adhesin fimh via finger trap-like sheet twisting. *Cell.* (2010) **141**:645–55. doi: 10.1016/j.cell.2010.03.038
10. Sauer MM, Jakob RP, Eras J, Baday S, Eris D, Navarra G, et al. Catch-bond mechanism of the bacterial adhesin FimH. *Nat Commun.* (2016) **7**:10738. doi: 10.1038/ncomms10738
11. Waldron TT, Springer TA. Transmission of allostery through the lectin domain in selectin-mediated cell adhesion. *Proc Natl Acad Sci USA.* (2009) **106**:85–90. doi: 10.1073/pnas.0810620105
12. Preston RC, Jakob RP, Binder FPC, Sager CP, Ernst B, Maier T. E-selectin ligand complexes adopt an extended high-affinity conformation. *J Mol Cell Biol.* (2016) **8**:62–72. doi: 10.1093/jmcb/mjv046
13. Pereverzev YV, Prezhdo OV, Forero M, Sokurenko EV, Thomas WE. The two-pathway model for the catch-slip transition in biological adhesion. *Biophys J.* (2005) **89**:1446–54. doi: 10.1529/biophysj.105.062158
14. Evans E, Leung A, Heinrich V, Zhu C. Mechanical switching and coupling between two dissociation pathways in a P-selectin adhesion bond. *Proc Natl Acad Sci USA.* (2004) **101**:11281–6. doi: 10.1073/pnas.0401870101
15. Barsegov V, Thirumalai D. Dynamics of unbinding of cell adhesion molecules: Transition from catch to slip bonds. *Proc Natl Acad Sci USA.* (2005) **102**:1835–9. doi: 10.1073/pnas.0406938102
16. Ley K, Laudanna C, Cybulsky MI, Nourshargh S. Getting to the site of inflammation: the leukocyte adhesion cascade updated. *Nat Rev Immunol.* (2007) **7**:678–89. doi: 10.1038/nri2156
17. Zhang T, Mbanga BL, Yashin VV, Balazs AC. Tailoring the mechanical properties of nanoparticle networks that encompass biomimetic catch bonds. *J Polym Sci Part B.* (2018) **56**:105–18. doi: 10.1002/polb.24542
18. Iyer BVS, Yashin VV, Balazs AC. Harnessing biomimetic catch bonds to create mechanically robust nanoparticle networks. *Polymer.* (2015) **69**:310–20. doi: 10.1016/j.polymer.2015.01.015
19. Bell G. Models for the specific adhesion of cells to cells. *Science.* (1978) **200**:618–27. doi: 10.1126/science.347575
20. Ju L, Dong JF, Cruz MA, Zhu C. The N-terminal flanking region of the A1 domain regulates the force-dependent binding of von Willebrand factor to platelet glycoprotein Ib alpha. *J Biol Chem.* (2013) **288**:32289–301. doi: 10.1074/jbc.M113.504001
21. Evans E, Ritchie K. Strength of a weak bond connecting flexible polymer chains. *Biophys J.* (1999) **76**:2439–47. doi: 10.1016/S0006-3495(99)77399-6
22. Gillespie D. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys.* (1976) **22**:403–34. doi: 10.1016/0021-9991(76)90041-3
23. Sprakel J, Lindstroem SB, Kodger TE, Weitz DA. Stress enhancement in the delayed yielding of colloidal gels. *Phys Rev Lett.* (2011) **106**:248303. doi: 10.1103/PhysRevLett.106.248303
24. Hosono N, Kushner AM, Chung J, Palmans ARA, Guan Z, Meijer EW. Forced unfolding of single-chain polymeric nanoparticles. *J Am Chem Soc.* (2015) **137**:6880–8. doi: 10.1021/jacs.5b02967
25. Corbin P, Zimmerman S. Self-association without regard to prototropy. A heterocycle that forms extremely stable quadruply hydrogen-bonded dimers. *J Am Chem Soc.* (1998) **120**:9710–1. doi: 10.1021/ja981884d
26. Sakai H, Sumi T, Aoki D, Goseki R, Otsuka H. Thermally stable radical-type mechanochromic polymers based on difluorenylsuccinonitrile. *ACS Macro Lett.* (2018) **7**:1359–63. doi: 10.1021/acsmacrolett.8b00755
27. Kato S, Ishizuki K, Aoki D, Goseki R, Otsuka H. Freezing-induced mechanoluminescence of polymer gels. *ACS Macro Lett.* (2018) **7**:1087–91. doi: 10.1021/acsmacrolett.8b00521
28. Gossweiler GR, Kouznetsova TB, Craig SL. Force-rate characterization of two spiropyran-based molecular force probes. *J Am Chem Soc.* (2015) **137**:6148–51. doi: 10.1021/jacs.5b02492
29. Davis DA, Hamilton A, Yang J, Cremar LD, Van Gough D, Potisek SL, et al. Force-induced activation of covalent bonds in mechanoresponsive polymeric materials. *Nature.* (2009) **459**:68–72. doi: 10.1038/nature07970
30. Achilleos DS, Hatton TA, Vamvakaki M. Light-regulated supramolecular engineering of polymeric nanocapsules. *J Am Chem Soc.* (2012) **134**:5726–9. doi: 10.1021/ja212177q
31. Zhang L, Dai L, Rong Y, Liu Z, Tong D, Huang Y, et al. Light-triggered reversible self-assembly of gold nanoparticle oligomers for tunable SERS. *Langmuir.* (2015) **31**:1164–71. doi: 10.1021/la504365b
32. Gostl R, Sijbesma RP. pi-extended anthracenes as sensitive probes for mechanical stress. *Chem Sci.* (2016) **7**:370–5. doi: 10.1039/C5SC03297K

33. Klukovich HM, Kouznetsova TB, Kean ZS, Lenhardt JM, Craig SL. A backbone lever-arm effect enhances polymer mechanochemistry. *Nat Chem.* (2013) **5**:110–4. doi: 10.1038/nchem.1540

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

SPACE PHYSICS

Raphael Marschall is a young space scientist working on small bodies in the solar system, comets and asteroids. He started his career at the University of Bern with a Bachelor and a Master thesis, both from the Institute of Theoretical Physics. Then he moved to the group of Nicolas Thomas to work on the observations of comet 67P/Churyumov-Gerasimenko from the Rosetta mission, on which he completed his PhD in 2017. The Faculty of Science at the University of Bern selected this as the best PhD thesis in physics that year. After a first 1.5-year postdoc at the International Space Science Institute in Bern he moved to the Southwest Research Institute in Boulder with a grant from the Swiss National Science Foundation. Recently he became a Postdoctoral Researcher at SwRI, where he works primarily on Trojan asteroids. He has some 15 publications with more than 700 citations.

The paper by Marschall et al. addresses the problem of dust emission from comets. The Rosetta mission to comet 67P/Churyumov–Gerasimenko has revealed that potentially a significant fraction of dust emitted from the comet will fall back to its surface. Knowing the total mass loss (volatile + dust) from the comet as well as the volatile mass loss (using gas dynamics models) can therefore fundamentally not determine the dust mass emitted from the surface. This is because only a part of the emitted dust escapes the nucleus' gravity contributing to the total mass loss while the remaining fraction falls back onto the surface. To solve this problem the authors have developed and applied state of the art 3D gas and dust dynamics models to simultaneously constrain the gas and dust emission. The models have been constrained using the ROSINA mass-spectrometer data for the gas and the OSIRIS imaging data for the dust. This modelling approach allows to not only constrain the gas and dust production rates but also simultaneously the dust size distribution, the dust-to-gas ratio, and the fraction of dust falling back to the comet surface. The authors found that the dust-to-gas ratio of comet 67P is of the order of unity and that the comet emitted about 5 megatons of dust during its 2015 apparition. The most likely dust size distribution was found to have a differential power law slope of -3.7, and the fraction of dust falling back to the surface is of the order of 10%. This corresponds to roughly 10 cm in deposition on the smooth northern planes. Finally, it was found that the smallest dust grain size must be strictly smaller than 30 microns.

frontiers in Physics

ORIGINAL RESEARCH
published: 24 June 2020
doi: 10.3389/fphy.2020.00227

The Dust-to-Gas Ratio, Size Distribution, and Dust Fall-Back Fraction of Comet 67P/Churyumov-Gerasimenko: Inferences From Linking the Optical and Dynamical Properties of the Inner Comae

Raphael Marschall[1]*, *Johannes Markkanen*[2], *Selina-Barbara Gerig*[3,4], *Olga Pinzón-Rodríguez*[3], *Nicolas Thomas*[3,4] and *Jong-Shinn Wu*[5]

[1] Department of Space Studies, Southwest Research Institute, Boulder, CO, United States, [2] Department Planets and Comets, Max Planck Institute for Solar System Research, Göttingen, Germany, [3] Space Research and Planetary Sciences, Physikalisches Institut, Universität Bern, Bern, Switzerland, [4] NCCR PlanetS, Bern, Switzerland, [5] Department of Mechanical Engineering, National Chiao Tung University, Hsinchu, Taiwan

OPEN ACCESS

Edited by:
Luca Sorriso-Valvo,
National Research Council, Italy

Reviewed by:
Nickolay Ivchenko,
Royal Institute of Technology, Sweden
ZhongYi Lin,
National Central University, Taiwan

***Correspondence:**
Raphael Marschall
marschall@boulder.swri.edu

Specialty section:
This article was submitted to Space Physics, a section of the journal Frontiers in Physics

Received: 09 November 2019
Accepted: 26 May 2020
Published: 24 June 2020

Citation:
Marschall R, Markkanen J, Gerig S-B, Pinzón-Rodríguez O, Thomas N and Wu J-S (2020) The Dust-to-Gas Ratio, Size Distribution, and Dust Fall-Back Fraction of Comet 67P/Churyumov-Gerasimenko: Inferences From Linking the Optical and Dynamical Properties of the Inner Comae. Front. Phys. 8:227. doi: 10.3389/fphy.2020.00227

In this work, we present results that simultaneously constrain the dust size distribution, dust-to-gas ratio, fraction of dust re-deposition, and total mass production rates for comet 67P/Churyumov-Gerasimenko. We use a 3D Direct Simulation Monte Carlo (DSMC) gas dynamics code to simulate the inner gas coma of the comet for the duration of the Rosetta mission. The gas model is constrained by ROSINA/COPS data. Further, we simulate for different epochs the inner dust coma using a 3D dust dynamics code including gas drag and the nucleus' gravity. Using advanced dust scattering properties these results are used to produce synthetic images that can be compared to the OSIRIS data set. These simulations allow us to constrain the properties of the dust coma and the total gas and dust production rates. We determined a total volatile mass loss of $(6.1 \pm 1.5) \cdot 10^9$ kg during the 2015 apparition. Further, we found that power-laws with $q = 3.7^{+0.57}_{-0.078}$ are consistent with the data. This results in a total of $5.1^{+6.0}_{-4.9} \cdot 10^9$ kg of dust being ejected from the nucleus surface, of which $4.4^{+4.9}_{-4.2} \cdot 10^9$ kg escape to space and $6.8^{+11}_{-6.8} \cdot 10^8$ kg (or an equivalent of 14^{+22}_{-14} cm over the smooth regions) is re-deposited on the surface. This leads to a dust-to-gas ratio of $0.73^{+1.3}_{-0.70}$ for the escaping material and $0.84^{+1.6}_{-0.81}$ for the ejected material. We have further found that the smallest dust size must be strictly smaller than $\sim 30\mu$m and nominally even smaller than $\sim 12\mu$m.

Keywords: comets, coma, 67P/Churyumov-Gerasimenko, dust-to-gas ratio, size distribution, modeling, dust dynamics

1. INTRODUCTION

The European Space Agency's (ESA) Rosetta mission escorted comet 67P/Churyumov-Gerasimenko (hereafter 67P) from August 2014 to September 2016 along its orbit through the inner Solar System. It watched as the comet's activity started to develop at large heliocentric distances, come to its culmination at perihelion, and decline as the comet

traveled out toward Jupiter's orbit. This long-term continuous monitoring of the comet's activity has provided an unprecedented wealth of data on this comet and its activity.

The observations revealed a complex bi-lobate shape [1, 2] and diverse morphology [3]. As a comet approaches the Sun it is heated and the ices start sublimating and ripping with them dust particles. Thus one of the important questions to be answered was what the bulk of the comet was made of i.e., what the bulk refractory-to-volatile ratio is. In the simplified view where any ejected material is lost to space two measurements are sufficient to determine this ratio. First, the total mass loss during one apparition measured by the Radio Science Investigation (RSI) [4]. Second, the total volatile mass loss which can be indirectly determined by the *in-situ* measurements of the gas density (e.g., [5–7]) or remote sensing data (e.g., [8–11]). In this simple case, the refractory-to-volatile ratio can be immediately inferred from those two measurements. But the complex surface morphology has revealed large dust deposits [12] that indicate that possibly a large fraction of the ejected dust is re-deposited [13]. If that is indeed the case, then the two above mentioned quantities cannot constrain the total dust mass ejected but rather only the dust mass escaping the nucleus gravity. Further, the process of dust fall-back obscures the emitted dust-to-gas ratio.

One way of constraining the amount of fall-back material would be to attempt to measure the actual change in elevation of the surface as a function of time from local or global digital terrain models (DTM). We cannot assess at this point if that is indeed feasible with the Optical, Spectroscopic and Infrared Remote Imaging System (OSIRIS, [14]) data set. Another way is to couple the scattering properties of the dust with a dynamical model of the dust coma constrained by the brightness of the dust coma. In this work, we have adopted the latter approach and modeled the inner gas and dust comae for the entire Rosetta mission. We use Rosetta Orbiter Spectrometer for Ion and Neutral Analysis (ROSINA, [15]) data to constrain the gas production rate and OSIRIS data for our dust models. To constrain the dust models we compare the dust coma brightness as measured by OSIRIS to synthetic model images. This process links several dust parameters that are otherwise not easily combined. In particular, we will show how the dust size distribution, the dust-to-gas ratio, the fraction of fall-back and the optical properties are inter-dependant and thus cannot be determined independently.

In section 2, we will describe the method used and lay out the assumptions we have made. Furthermore, we will point out the free parameters of the models, that need constraining through Rosetta data. Some theoretical considerations are presented in section 3. We will discuss the results of our work in section 4 and summarize and conclude our work in section 5.

2. METHOD

In this work, we have used the modeling approach (and in particular our DRAG3D model for the dust coma) described in detail in Marschall et al. [16]. This approach has been successfully applied for the analysis and interpretation of multiple Rosetta instruments, in particular ROSINA, MIRO (Microwave Instrument for the Rosetta Orbiter), VIRTIS (Visible and Infrared Thermal Imaging Spectrometer), and OSIRIS [16–19]. While in previous work we have applied this approach to specific epochs of the Rosetta mission, we have employed it here to cover the entire mission period to study longer-term processes.

In the following, we will briefly repeat some of the most important parts of the modeling elements and refer to Marschall et al. [16] for a detailed description.

2.1. General Assumptions

The calculation of the 3D gas flow field using the Direct Simulation Monte Carlo (DSMC) method is very computationally expensive and it is therefore currently not feasible to cover the entire escort phase of Rosetta (from August 2014 to September 2016) with a high temporal resolution. It is thus necessary to split the comet's orbit into a number of epochs that are computationally feasible and then interpolate between the results using a linear scaling between epochs. To ensure that the calculated results are representative of the respective epoch we make sure that during each of the epochs neither the total solar energy reaching the surface nor where the energy strikes the surface changes substantially. The amount of energy deposited is driven primarily by the heliocentric distance, R_h, while the location of deposition apart from the rotation of the comet is controlled by the sub-solar latitude, *LAT*. We thus chose that the inverse square of the heliocentric distance of the comet's location at the start and end time of each epoch shall be within 15% of the location at the center date of each epoch. Furthermore that the difference in sub-solar latitude be less than 5° from the center time of epoch to the start and end of the epoch, respectively. This leads to the 20 epochs listed in **Table 1** and illustrated in **Figure 1**. Simulations were run for the center time of each epoch. This choice also ensures that we cover the exact dates of the in- and outbound equinox (epochs 6 and 18) as well as perihelion (epoch 11) and summer solstice (epoch 12).

The basis of all simulations is the 3D shape model by Preusker et al. [2]. We use a decimated model with $\sim 440'000$ facets due to our computational constraints. To fully define the illumination condition we need to select the sub-solar longitude in addition to the heliocentric distance and sub-solar latitude which are set by the choice of epoch. For each epoch we have run simulations for sub-solar longitudes of 0°, 30°, 60°, 90°, 120°, 150°, 180°, 210°, 240°, 270°, 300°, and 330°. This results in a total of 240 different illumination conditions for the entire mission period.

For each illumination condition, we calculate the incidence angle (angle between the surface normal and the direction of the Sun) of each facet taking into account self-shadowing. This allows calculating the solar energy entering the surface neglecting re-radiation from other facets. By means of a simple energy balance of the incoming solar energy, thermal re-radiation and sublimation we can calculate the sublimation temperature and the sublimation rate of each facet assuming pure water ice.

We do not take into account any emission from shadowed facets, be it due to local night or mutual shadowing by other parts of the nucleus. The calculated pure ice sublimation rate of each facet needs to be scaled to match observed sublimation rates at

TABLE 1 | Start, center, and end time for each of the epochs as well as the heliocentric distance and sub-solar latitude of the center time of each epoch.

Epoch	Start time	Center time	End time	R_h [AU]	*LAT* [°]	Qualifier
1	2014-07-05 18H	2014-08-26 12H	2014-10-03 18H	3.4703	43.3	
2	2014-10-03 18H	2014-11-11 12H	2014-12-10 18H	2.9844	37.1	
3	2014-12-10 18H	2015-01-10 00H	2015-02-02 06H	2.5667	29.8	
4	2015-02-02 06H	2015-02-26 18H	2015-03-18 00H	2.2083	21.1	
5	2015-03-18 00H	2015-04-05 00H	2015-04-20 12H	1.9213	11.6	
6	2015-04-20 12H	2015-05-04 06H	2015-05-16 12H	1.7006	1.6	Inbound equinox
7	2015-05-16 12H	2015-05-27 12H	2015-06-06 18H	1.5372	-8.3	
8	2015-06-06 18H	2015-06-16 06H	2015-06-25 06H	1.4156	-18.2	
9	2015-06-25 06H	2015-07-04 00H	2015-07-12 12H	1.3278	-28.1	
10	2015-07-12 12H	2015-07-21 06H	2015-07-30 18H	1.2695	-38.0	
11	2015-07-30 18H	2015-08-11 00H	2015-08-21 12H	1.2432	-47.9	Perihelion
12	2015-08-21 12H	2015-09-02 18H	2015-09-16 12H	1.2742	-52.3	Summer solstice
13	2015-09-16 12H	2015-09-27 12H	2015-10-12 00H	1.3709	-48.0	
14	2015-10-12 00H	2015-10-25 06H	2015-11-07 18H	1.5344	-38.0	
15	2015-11-07 18H	2015-11-22 00H	2015-12-07 12H	1.7307	-28.1	
16	2015-12-07 12H	2015-12-24 12H	2016-01-12 18H	1.9778	-18.2	
17	2016-01-12 18H	2016-02-01 18H	2016-02-27 06H	2.2796	-8.6	
18	2016-02-27 06H	2016-03-22 12H	2016-04-23 06H	2.6491	0.4	Outbound equinox
19	2016-04-23 06H	2016-05-24 00H	2016-07-04 00H	3.0797	8.9	
20	2016-07-04 00H	2016-08-13 18H	2016-09-28 00H	3.5819	17.0	

All times are given in the format YYYY-MM-DD hhH UTC.

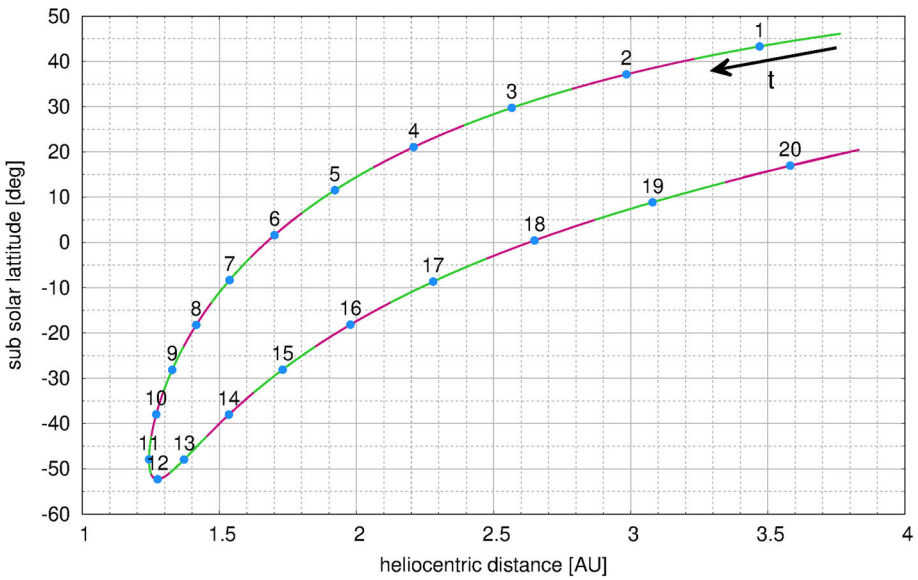


FIGURE 1 | Heliocentric distance vs. sub-solar latitude of comet 67P during the escort phase of Rosetta as well as epochs used in this work.

67P. Here we assume a pure H_2O ice surface that is areally mixed with inert refractory surface akin to a checkerboard pattern. This surface fraction of the facet covered by ice, which is a priori not known, is a free parameter of the model. We refer to this scaling factor as the *effective active fraction* (EAF). This factor only has a physical interpretation for a pure ice surface where it would represents the fraction of pure ice of an areally mixed surface needed for a specific sublimation flux. In general though it is not a physical parameter and should not be interpreted as such.

In the next steps, we calculate the gas and dust flow fields in three dimensions. We then perform a column integration along the line-of-sight through the dust coma for a specific viewing geometry of the OSIRIS NAC (narrow-angle) and WAC (wide-angle) cameras [14] and convolve the dust column densities with

the optical properties of the dust to arrive at absolute radiance values that can be compared with the OSIRIS images. One major assumption that goes into this approach is that there is no significant back-coupling from the dust to the gas allowing a sequential treatment of the two flows. For low dust-to-gas mass ratios, this is certainly justified [16] but will break down when a lot of dust is released. We will further discuss this limitation later on.

2.2. Gas Kinetic Simulations

The gas flow-field is calculated using the DSMC technique. The code used is called `UltraSPARTS`[1] and is a commercialized derivative of the `PDSC++` code [20] used in previous papers (e.g., [16, 17]). `PDSC++` is a C++ based, parallel DSMC code which is capable of simulating 2D, 2D-axisymmetric, and 3D flow fields. The code has been developed over the past 15 years [21–23] and contains several important features including the implementation of 2D and 3D hybrid unstructured grids, a transient adaptive sub-cell method (TAS) for denser flows, and a variable time-step scheme (VTS). In the parallel version, computational tasks are distributed using the Message Passing Interface (MPI) protocol. The improved `UltraSPARTS` (Ultra-fast Statistical PARTicle Simulation Package) has been applied to 67P [18, 19]. Here we simulate the full 3D gas flow up to a distance of 10 km from the nucleus center.

The sublimation temperature and flux—calculated as described above—for each facet are set as initial conditions of the simulation. This includes implicitly the assumption of the appropriate EAF. We assume here that the EAF of all facets are the same (i.e., homogeneous surface properties) but can change from epoch to epoch. This results in one value for the global EAF per epoch. Though we know from previous works (e.g., [5, 16, 24, 25]) that there are regional inhomogeneities that can be encoded in EAF it is not the focus of this work to constrain these inhomogeneities. Rather we seek a global estimate of the fluxes and dynamical behaviors. Because the EAF is a free parameter it needs to be constrained by data. In our case, we determine the EAF by comparing modeled densities extrapolated to the Rosetta position and actual COmet Pressure Sensor (COPS; [15]) measurements during each epoch. Within each epoch where we match the sub-solar longitude of a measurement, we extract from the respective simulation the gas number density at the position of the spacecraft. If the spacecraft distance is larger than 10 km we extrapolate the value from the 10 km surface to the spacecraft distance assuming free radial outflow. This assumption is well-justified as shown in Marschall et al. [16]. Though this does not capture the detailed structure of the ROSINA/COPS data it does account accurately for the average activity level at each epoch. **Table 2** shows the EAF used and the resulting average global H_2O production rate for one comet day.

The global gas production rate as a function of time is shown in **Figure 11** (purple band in top panel). Because we have used the total ROSINA/COPS data—which also contains the other gas species other than water—to constrain our emission, these values should be understood as a proxy for the entire emission. We have interpolated between the epochs using a local second-order polynomial. The fit for each epoch, i, includes three epochs i−1, i, i+1. The fitting parameters to calculate the mean gas production rate as a function of ephemeris time (ET) is shown in **Table 3**. The resulting integrated mass loss over the shown period adds up to

[1]http://www.plasmati.tw

TABLE 2 | The effective active fraction (EAF) assumed and the resulting average H_2O production rate for each epoch.

Epoch	EAF	Q_{H_2O} [kg s^{-1}]
1	0.87	1.047
2	1.40	3.120
3	1.60	6.178
4	2.10	12.42
5	2.80	24.74
6	3.24	38.54
7	6.00	92.38
8	8.80	168.2
9	10.9	249.9
10	12.5	328.9
11	16.5	473.2
12	30.1	827.0
13	17.3	393.3
14	11.4	192.7
15	9.14	110.1
16	6.84	55.98
17	2.61	13.92
18	1.33	4.321
19	0.32	0.592
20	0.41	0.393

TABLE 3 | The mean gas production rate, $\bar{q}_g$ [kg s^{-1}] as a function of ephemeris time (ET): $\bar{q}_g(ET) = a \cdot ET^2 + b \cdot ET + c$.

Epoch	a	b	c
2	2.400340e-14	-2.204283e-05	5.061404e+03
3	9.908836e-14	-9.285540e-05	2.175673e+04
4	3.151072e-13	-2.985843e-04	7.073810e+04
5	2.843195e-13	-2.690371e-04	6.364899e+04
6	4.705372e-12	-4.537366e-03	1.093862e+06
7	4.739402e-12	-4.570374e-03	1.101867e+06
8	2.736443e-12	-2.620080e-03	6.271145e+05
9	-8.310099e-14	1.344634e-04	-4.564429e+04
10	8.356934e-12	-8.136523e-03	1.980682e+06
11	2.648205e-11	-2.595811e-02	6.361453e+06
12	-9.326697e-11	9.223562e-02	-2.280309e+07
13	2.625803e-11	-2.622753e-02	6.549455e+06
14	1.025950e-11	-1.029856e-02	2.584549e+06
15	2.918884e-12	-2.954660e-03	7.477641e+05
16	1.108671e-12	-1.134214e-03	2.900836e+05
17	1.322669e-12	-1.350748e-03	3.448580e+05
18	1.589655e-13	-1.643041e-04	4.245606e+04
19	5.313567e-14	-5.537983e-05	1.442952e+04

$(6.1 \pm 1.5) \cdot 10^9$ kg. This is well in line with values published in other works as e.g., $(6.3 \pm 2.0) \cdot 10^9$ kg [7] or $(5.8 \pm 1.8) \cdot 10^9$ kg [6]. The error arises from the uncertainty of the data (up to 15%; [26]) although the relative errors are probably smaller (M. Rubin, pers. comm.) and our model (5–10%; [27]) as well as from the scatter from the comparison of the data and our model. As it was not the goal of this work to constrain as precisely as possible the surface-emission distribution it is nevertheless noteworthy that our estimates come so close to the other published values. This illustrates that it is not necessary to know the surface-emission distribution well to estimate the total global volatile loss. Rather simple assumptions of the surface response is sufficient for such an estimate. Though it is not that surprising, because as pointed out in Marschall et al. [28] the global gas production rate can be fairly well-estimated by even simplified models. Our peak production rate is reached at the summer solstice (epoch 12) and not perihelion (epoch 11) and therefore roughly 22 days post-perihelion. This is in line with dust coma measurements by OSIRIS. Gerig et al. [18] reported peak dust coma brightness 20 days post-perihelion. This also hints at the fact that the obliquity plays an important role in the activity of comets. Though the heliocentric distance still is the main driver of the gas and dust activity ($O(1)$) it is the obliquity/season that controls the second order. The coincidence of the peak gas activity with the peak dust activity also indicates that the dust activity is mainly driven by H_2O or at least near-surface volatiles without a significant thermal lag.

2.3. Dust Dynamic Simulations

After the gas flow field has been evaluated, we calculate the dust flow field by injecting dust test particles into the flow. We use a typical approach for computing the dust motion in a gas flow-field taking into account gas drag and the comet gravity using our DRAG3D dust coma model detailed in Marschall et al. [16] and the references therein. We assume that the dust mass production rate is proportional to the gas mass production rate and that the dust size distribution does not vary across the surface except in cases where certain dust sizes are no longer lifted because the gas pressure is too low to surpass the local gravity. The dust size distribution is thus only naturally modified by the dynamics and lifting process. It is assumed that the dust particles are at rest on the surface (i.e., the ejection velocity is 0). The dust-to-gas mass ratio as well as the dust size distribution at the surface are free parameters of the model and will be constrained by the data as described below. Due to the presence of gravity, large dust particles may not reach escape speed and eventually return to the surface. The flux of back-fall particles is thus a further output of the model. It is important to note that we assume that the dust particles are desiccated, i.e., contain no significant amounts of volatiles that evaporate while airborne. They may still be wet but do not outgas significantly. This is a consequence of our assumption that there is no significant back-coupling of the dust flow onto the gas flow.

For each epoch (except for two) we have selected one OSIRIS image where the illumination conditions of the image match one of the gas simulations (see **Table 1**). The images used in this work, as well as some camera and geometric properties, are shown in **Table 4**. Two main criteria were used to select these images. First, the images needed a large enough field of view such that projected distance (impact parameter, b) in the image plane from the center of the comet to the edge of the image at each side was at least 9 km. Why this is an important constraint will be described in the next paragraph. Second, images need a sufficient signal to noise such that the dust coma brightness could be measured well. These two constraints unfortunately, eliminated all images for epoch 2 and 20. Epoch 2 included the 10 km orbit phase and thus did not provide large enough fields-of-view while the signal-to-noise was bad in epoch 20 due to the very low activity of the comet. Most images we have used were taken by the wide-angle camera (WAC) and filter 18 (central wavelength, $\lambda_c = 612$ nm) and at cometocentric distances between 87 and 635 km and phase angles between 37° and 108°.

The dust field is calculated for each image using 41 different dust sizes from 10 nm to 1 m. The dust sizes are logarithmically spaced with five dust sizes per decade. The particles are assumed to be spherical and have a density of 533 kg m^{-3} matching roughly the bulk density of the nucleus [2]. Even though all dust sizes are simulated, not all of them contribute to the dust brightness in the coma. This is because the particles larger than a certain size might not all be lifted because the gas pressure cannot overcome gravity. Thus the number of dust sizes present in the coma depends on the heliocentric distance (epoch). The upper size limit (largest liftable size) is thus naturally determined and thus an outcome of the simulation. What the smallest dust size should be is unless. The smallest diameter of particle sub-units measured by MIDAS [29] is 100 nm. Whether these could also be the smallest dust particles in the coma or if these measurements have an *in-situ* collection bias at the spacecraft is not clear. One could imagine that very small particles might not have been collected because of spacecraft and dust charging. Della Corte et al. [30] showed that particles, for which the ratio of the particle charge to its kinetic energy entering the electrostatic field of the space craft $q/E_k > 0.24CJ^{-1}$, will not reach the spacecraft. We will therefore leave this issue open for the moment and examine the impact of the smallest size on the results in section 4.

Once the 3D dust field is simulated we calculate the dust column densities of each size for the specific viewing geometry of the respective OSIRIS image. The final image is composed by weighting the different dust sizes according to a specific dust size distribution and convolving the column densities with the scattering properties described in section 2.4.

For each of the images we compare the integrated radiance of the dust coma along an aperture with impact parameter $b = 11$ km and compare it to that of the synthetic images. Again it is not our goal in this work to match the structure of the inner dust coma but rather the overall global behavior. Gerig et al. [18] showed from OSIRIS data that the dust flow goes over to free radial outflow at an average impact parameter of $b \sim 11$ km. This is in line with theoretical considerations of dusty flows [31]. Beyond that point, the dust brightness falls off with $1/b$ as expected for a freely expanding radial flow. For that reason, we have chosen $b = 11$ km to be within the free-flow regime. If the field-of-view was not large enough we used the maximum available impact parameter.

TABLE 4 | List of OSIRIS images used in each of the epochs as well as their filter, central wavelength (λ_c), phase angle (α), and cometo-centric distance (D_{cc}).

Epoch	OSIRIS image name	Filter	λ_c [nm]	α [°]	D_{cc} [km]
1	WAC_2014-08-16T13.01.44.647	F18	612.6	37.85	100.03
2	-	-	-	-	-
3	WAC_2015-02-09T19.10.19.184	F18	612.6	87.43	105.61
4	WAC_2015-02-19T23.11.20.158	F18	612.6	73.64	136.20
5	WAC_2015-04-11T08.14.54.119	F18	612.6	88.32	140.96
6	WAC_2015-05-05T11.27.54.523	F18	612.6	64.68	165.48
7	WAC_2015-05-22T06.36.34.903	F18	612.6	59.82	134.89
8	WAC_2015-06-16T03.22.37.523	F18	612.6	89.06	220.19
9	WAC_2015-07-04T11.11.58.808	F18	612.6	89.83	176.63
10	WAC_2015-07-19T00.20.07.696	F18	612.6	89.25	181.28
11	WAC_2015-08-09T09.13.16.574	F18	612.6	89.09	306.41
12	WAC_2015-09-02T07.58.47.075	F18	612.6	72.76	414.13
13	WAC_2015-09-25T11.39.29.053	F18	612.6	56.35	635.77
14	NAC_2015-10-14T07.03.45.244	F22	649.2	64.45	497.73
15	WAC_2015-11-22T21.43.06.876	F18	612.6	89.23	127.75
16	WAC_2015-12-30T08.12.53.526	F18	612.6	89.58	87.682
17	WAC_2016-01-13T07.03.19.181	F18	612.6	88.05	87.974
18	WAC_2016-03-25T02.15.44.556	F18	612.6	108.2	270.11
19	WAC_2016-04-13T18.47.58.431	F18	612.6	76.41	106.43
20	-	-	-	-	

The name of the OSIRIS image contains the camera used (WAC, wide angle camera; NAC, narrow angle camera) and the time stamp in the format {YYYY-MM-DD}T{hh.mm.ss.ms}.

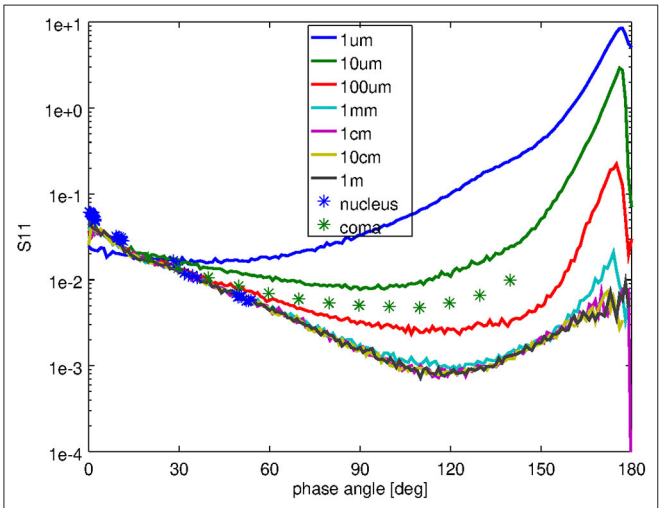


FIGURE 2 | Shows the phase functions, S_{11}, as a function of phase angle for different particle sizes (solid lines). The green stars represent the measurements of the coma phase function by Bertini et al. [38] while the blue stars show the nucleus phase function as measured by Fornasier et al. [39], Feller et al. [40], and Masoumzadeh et al. [41].

2.4. Scattering Model

Previously, we have used a spherical particle model and a Mie scattering code in our modeling pipeline. Here we use a much more sophisticated approach based on the recently introduced radiative transfer with reciprocal transactions framework [32, 33]. The approach allows for scattering analysis of large irregularly shaped particles with wavelength-sized details. Here, the dust particles are considered to be irregular aggregates composed of sub-micrometer-sized organic grains and micrometer-sized silicate grains. Such a particle model has been found to be in good agreement with OSIRIS [34] and VIRTIS [35] phase function measurements. The refractive index for silicate grains is assumed to be $m = 1.6048692 + i0.0015341$ corresponding to magnesium iron pyroxine [36] and for organic grains $m = 1.55950 + i0.42964$ corresponding to amorphous carbon [37]. At 612 nm (WAC filter 18) the resulting scattering phase functions for different particle sizes normalized to the geometric albedo are shown in **Figure 2**. The figure shows good agreement of the phase function of large particles ($>$ 1 cm) with the nucleus phase function as measured by OSIRIS [39–41]. This should indeed be the case because larger particles should behave more and more like "small comets" themselves and thus be representative of the nucleus scattering properties. For small particles, the best agreement of a single dust size with the coma phases function [38] is between 10 and 100 μm. The numerical method of Markkanen et al. [34] is not applicable to particles smaller than 1μm. Thus for the particles smaller than 1μm we use a Mie scattering code to determine the scattering properties [see 16] matching the single scattering albedo of the Mie result with the approach of Markkanen et al. [34] for 1μm particles. This gives us a smooth transition from the large particle region to the Rayleigh scattering region where particle's shape has a negligible effect on its scattering properties. This is a state of the art model and we have thus used its results throughout this work. But because the scattering model does have an effect on the results a re-evaluation of the results can be done if and when a better model arises.

3. THEORETICAL CONSIDERATION

To put some of our results in the next section (section 4) into context, we present first some general theoretical considerations of what we can expect, in particular with regards to the relationship between the dust-to-gas ratio and the dust size distribution. We thus consider first a simple model where the comet is represented as a sonic (i.e., the gas velocity near the surface is equal to the local sound velocity and defined by the thermodynamic properties of the gas and the surface temperature i.e., R_h) spherical source of ideal perfect gas (with specific heat ratio γ=1.33) accelerating spherical solid grains. The source shall have radius R_N, nucleus mass M_N, total gas production rate Q_g (kg/s). The motion of a spherical grain in a flow from such a source was studied in Zakharov et al. [31] for a wide range of conditions. They defined

$$\mathrm{Iv} = \frac{3Q_g}{32R_N r\rho_d \pi v_{g0}}, \tag{1}$$

and

$$\mathrm{Fu} = \frac{GM_N}{R_N}\frac{1}{(v_g^{max})^2} \tag{2}$$

which are dimensionless parameters, where ρ_d is the specific density of the dust particles, v_{g0} the gas velocity near the surface, v_g^{max} the theoretical maximal velocity of gas expansion (defined by the thermodynamic properties of the gas and the surface temperature i.e., R_h), and G the gravitational constant. *Iv* characterizes the ability of a dust particle to adjust to the gas velocity while *Fu* quantifies the importance of gravity. Zakharov et al. [31] found that for $Iv < 0.1$ (which is the case of 67P, and dust sizes $>$ 1 nm) the dust particles reach 90% of their terminal velocity at about $6 \cdot R_N$. The terminal velocity of particles with radius, r, varies as $v_d(r) \propto r^{-0.5}$ for small *Fu* (i.e., if gravity plays a minor role). The asymptotic dust velocities are given by:

$$v_d(r) = \left(\frac{\mathrm{Iv(r)}}{\mathrm{Iv(r_*)}}\right)^{1/2}, v_d(r_*) = \left(\frac{r}{r_*}\right)^{-1/2}, v_d(r_*) = r^{-1/2}C_{Iv} \tag{3}$$

where r_* and $v_d(r_*)$ are some referential size and corresponding terminal velocity, and C_{Iv} is a constant.

For a dust size distribution given by a power-law, r^{-q}, the normalized mass distribution, f_{md}, of particles ejected from the surface is

$$f_{md}(r) = \begin{cases} \frac{4-q}{r_{max}^{4-q}-r_{min}^{4-q}} r^{3-q}, & q \neq 4 \\ \ln\left(-\frac{r_{max}}{r_{min}}\right) r^{3-q}, & q = 4 \end{cases} \tag{4}$$

where r_{min} and r_{max} are the smallest and largest dust sizes ejected from the surface. In the following, we will not considered specially the case of $q = 4$. The dust production rate, Q_d, of each dust size is

$$Q_d(r) = \chi f_{md}(r) Q_g dr, \tag{5}$$

where $\chi = Q_d/Q_g$ is the total dust-to-gas mass loss rate. Therefore, the number density of dust particles with radius, r, at the radial distance, R, from the center of the nucleus is:

$$n(r,R) = \frac{\chi f_{md}(r) Q_g dr}{v_d(r) m_d(r) 4\pi R^2}. \tag{6}$$

The column density at the distance ϱ from the center of the nucleus in the image plane is:

$$n_{col}(r,\varrho) = \int_{-\infty}^{+\infty} n(r,R)dz = \frac{\chi f_{md}(r) Q_g dr}{v_d(r) m_d(r) 4}\frac{1}{\varrho}. \tag{7}$$

The total number of dust particles in a column within a circular observing aperture of radius $\Re$ is:

$$N(r,\Re) = \int_0^{\Re} n_{col}(r,\varrho) 2\pi\varrho d\varrho = \frac{\chi f_{md}(r) Q_g dr}{v_d(r) m_d(r) 2}\pi\Re. \tag{8}$$

The brightness is proportional to the flux F (W/m^2) gathered by an instrument which for an optically thin coma is:

$$F(r,\Re) = \mathcal{F} N(r,\Re)\pi r^2 q_{sca}(r)\frac{\varphi_{av}(r)}{4\pi}\frac{1}{\Delta^2} \tag{9}$$

where $\mathcal{F}$ is the incident flux, Δ is observational distance, q_{sca} is scattering efficiency and φ_{av} is the phase function averaged over phase angle. Substituting Equations (3), (4) and (8) in (9) we get:

$$F(r,\Re) = \frac{3}{32}\frac{\mathcal{F} Q_g \Re}{C_{Iv}\rho_d \Delta^2}\frac{4-q}{r_{max}^{4-q}-r_{min}^{4-q}} r^{\frac{5}{2}-q} \chi q_{sca}(r)\varphi_{av}(r) dr. \tag{10}$$

For fixed $\mathcal{F}$, Q_g, R_N, ρ_d, v_{g0}, r_*, $v_d(r_*)$, Δ, and $\Re$

$$F(r,\Re) = C\frac{4-q}{r_{max}^{4-q}-r_{min}^{4-q}} \chi r^{\frac{5}{2}-q} q_{sca}(r)\varphi_{av}(r) dr \tag{11}$$

where $C = \frac{3}{32}\frac{\mathcal{F} Q_g \Re}{C_{Iv}\rho_d \Delta^2}$ is a constant (for the given observational conditions).

For the optical properties, we make some simplifying assumption. First, we approximate the scattering efficiency $q_{sca}(r)$ to be:

$$q_{sca}(r) = \begin{cases} 0.233 \cdot 10^{24} \cdot r^{7/2}, & 10^{-8} \leq r < 2 \cdot 10^{-7} \\ 4.993 \cdot 10^{-5} \cdot r^{-2/3}, & 2 \cdot 10^{-7} \leq r < 2 \cdot 10^{-4} \\ 0.02, & 2 \cdot 10^{-4} \leq r \leq 1.0 \end{cases} \tag{12}$$

Figure 3 shows the computed q_{sca} from section 2.4, the fitted q_{fit} scattering efficiency and the relative difference. In this fit, we used "round numbers" (i.e., this is a very rough fit). This simplification results in differences of $<$ 50%. For the phase function φ_{av} we assume it to be constant for all sizes. We estimate an error of the order of a factor of 2 from this simplification.

Under the assumption we made the integration of Equation (11) becomes trivial.

For a given gas production rate Q_g the maximum dust size a_{max} is also constant. **Figure 4** shows how the dust-to-gas mass loss rate Q_d/Q_g varies as a function or the power-law exponent of

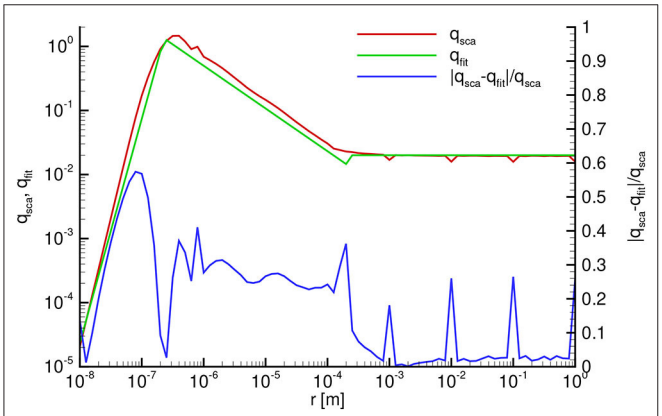


FIGURE 3 | Scattering efficiency as a function of dust radius: computed (comp, red) according to section 2.4, and fitted (fit, green). The relative difference between the computed and fitted is shown in blue.

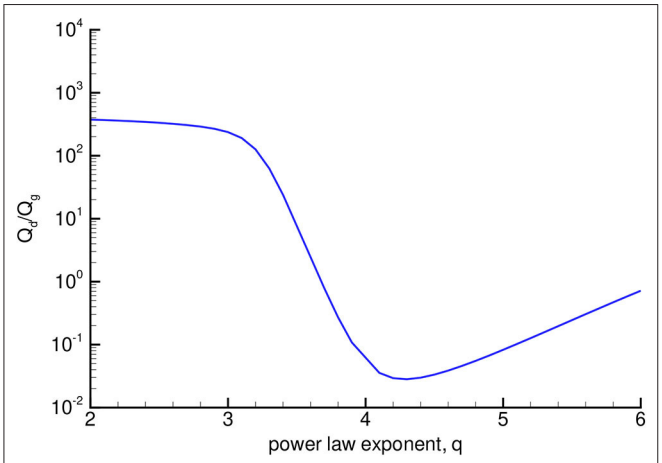


FIGURE 4 | Dust-to-gas mass production rate ratio vs. power-law exponent for constant dust brightness.

the dust size distribution for the same brightness. With increasing power-law exponent q from minimal value to $\approx$ 3 the Q_d/Q_g is slowly decreasing (since in this case practically all dust mass is concentrating in a narrow range of largest sizes), but with increasing q from 3 to 4 the ratio Q_d/Q_g is strongly decreasing. For $4.5 < q < 6$ Q_d/Q_g is increasing (within one order of magnitude). This inflection point of Q_d/Q_g occurs at the transition from dust grains distribution with most mass being in the large dust sizes to where most mass is in the small dust sizes.

The growth of Q_d/Q_g for $q > 4.5$ is a consequence of the strong decrease of q_{sca} for small sizes, therefore, in order to maintain the same brightness, it is necessary to eject more dust.

We should remember that this analytical result (**Figure 4**) assumed several important simplifications:

1. We assumed that the dust expansion is strictly radial;
2. For evaluation of the dust brightness we used simplified optical properties (e.g., isotropic phase function);
3. We assumed that gravity plays only a minor role;
4. We assumed that the dust does not affect the gas flow.

We will discuss in the next section how these assumptions [in particular (1) and (3)] change the result.

4. RESULTS AND DISCUSSION

To convolve the results of the dust dynamics model with the scattering properties to arrive at synthetic OSIRIS images we need to assume a dust size distribution. As is commonly done we presume that the number of particles, n, of a certain radius, r, follows a power-law:

$$n(r) \propto r^{-q} \quad , \tag{13}$$

where q is the differential power-law exponent. **Figure 5** illustrates an example of an OSIRIS and synthetic image for epoch 12 (solstice). As described in section 2.3, we extract the integrated brightness along a circle with a constant impact parameter of $b = 11$ km where possible (illustrated in the figure with the red circles). We should stress here again that it was not the aim of this work to match the emission distribution on the surface and thus all the structures in the coma.

For a given gas production rate, the three major factors controlling the brightness (see Equation (11) for more detail) of the dust coma are:

1. The dust-to-gas mass production rate ratio, Q_d/Q_g, at the surface;
2. The dust size distribution (i.e., the power-law exponent, q) at the surface;
3. The scattering properties of the dust particles.

We should note that although we assume a uniform surface (i.e., globally constant Q_d/Q_g and q) the actual values at each facet vary depending on the local gas flux. If a particular facets' local flux is too low to lift a certain particle size the resulting dust flux and size distribution from that surface facet will differ locally from the nominal values. The three input parameters above are not known a priori and are thus initially free parameters and in need of constraints. We have fixed the scattering properties by using published results that fit OSIRIS data [see sections 2.4 and 34]. This reduces the above parameters from three to two.

There are three further quantities that influence the coma brightness, but as we will show below their influence is small compared to the ones mentioned above. These are:

1. The smallest dust size, r_{min};
2. The largest dust size, r_{max};
3. The bulk density of the dust, ρ_d.

Of these three we will explore the influence of r_{min} and ρ_d on our results. We will not be artificially truncate the size range at large sizes by varying r_{max}. On the contrary, r_{max} will be naturally regulated due to the balance of forces at the surface. If a given local gas flux is not sufficient for lifting a certain size, that will determine the largest dust size from that surface element.

Apart from the parameters that directly influence the brightness, several indirect factors further constrain the curves Q_d/Q_g and q. We will discuss these constraints at the end of this section.

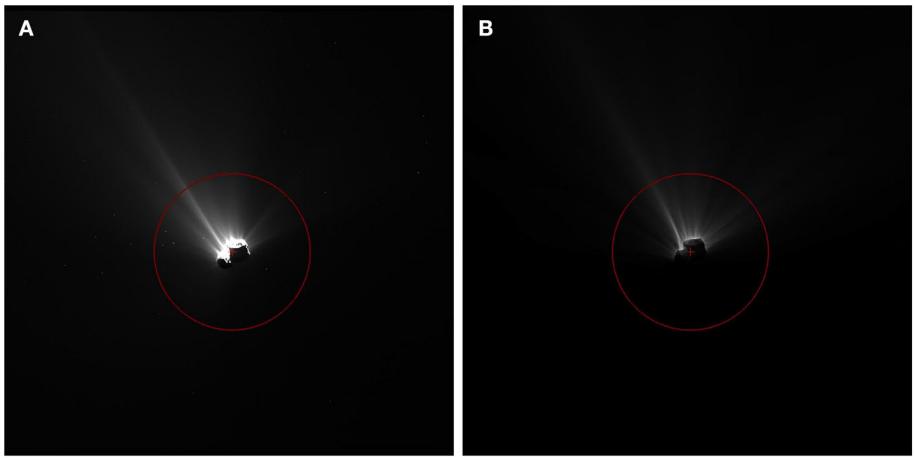


FIGURE 5 | Panel **(A)** shows the OSIRIS image WAC_2015-08-09T09.13.16.574Z of epoch 12 (solstice) with an enhanced contrast to show the dust coma. Panel **(B)** shows the synthetic image with power-law exponent of $q = 4$ of our dust model. The crosses in each panel indicate the center of the nucleus and the red circles indicate an impact parameter distance of $b = 11$ km.

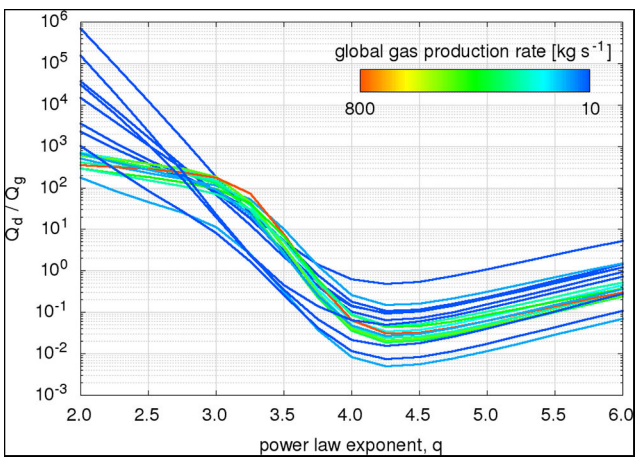


FIGURE 6 | The dust-to-gas mass production rate ratio, Q_d/Q_g as a function of the power-law exponent, q, is shown for each of the 18 OSIRIS images of this study. Each line represents an equal brightness curve where the respective q and Q_d/Q_g result in a synthetic image that matches the OSIRIS brightness. The colors of the curves indicate the global gas production rate of the respective epoch.

As has already been shown in Figure 12 of Marschall et al. [16] Q_d/Q_g and q are not independent. Knowing the brightness of the coma these two parameters constrain each other to a limiting set of parameter pairs. **Figure 6** shows Q_d/Q_g as a function of q for each of the 18 OSIRIS images of this study. Each line represents an equal brightness curve where the respective q and Q_d/Q_g result in a synthetic image that matches the OSIRIS brightness. Several things are noteworthy. First, all curves show minima between $q = 4$ and $q = 4.5$ and thus illustrate the inherent degeneracy between Q_d/Q_g and q. Second, all cases with shallow power-laws ($q < 3$) require very large Q_d/Q_g of at least 10 but in most cases around 100. Third, steep power-laws ($4 < q < 6$) in all but one case require much less dust mass, i.e., $Q_d/Q_g \leq 1$ to match the brightness of OSIRIS. Fourth, there is a clear trend in the gas production rate. As the gas production rate increases the slope in Q_d/Q_g for shallow power-laws ($q < 3$) becomes shallow, too. Or conversely for low gas production rates very high Q_d/Q_g are needed to match the OSIRIS brightness when the power-law is shallow. This has to do with the amount of dust that can be lifted and escape the nucleus' gravity.

Comparing **Figure 6** to the analytical solution presented in **Figure 4** of section 3 we see that for high gas production rates the model follows the analytical solution rather well. The places where we deviate from the analytical solution illustrate the effect of different physical processes. For the analytical solution we have assumed a minor (but not negligible) role of gravity. The effect of gravity can be seen in the low gas production rate cases with shallow power-laws. There, in contrast to the analytical model which levels off at smaller power-law exponents, the dust coma model results in ever higher Q_d/Q_g. This is caused by the inability to lift large particles from the entire surface and therefore a higher Q_d/Q_g is required to maintain the brightness. Thus, the deviations at low gas production rates and shallow power-laws exhibit the non-minor role of the nucleus' gravity. As in the analytical model for steeper size distributions most mass is in the smallest particles, which are weakly scattering and thus hardly contribute to the brightness. This is compensated by an increase of Q_d/Q_g at these steep power-laws.

Compared to previous work presented in Figure 12 of Marschall et al. [16] the Q_d/Q_g values we find here (in particular for $q < 3.5$) are much higher while the behavior of the curves for steeper power-laws is within the expected range. The two main reasons we find larger values at shallow slopes are: (1) Marschall et al. [16] assumed the scattering properties of astronomical silicate [42] which is much brighter than we now know; (2) we consider here considerably larger dust sizes as our upper limit. This extension of the size domain increases the Q_d/Q_g by orders of magnitude because of high fall back fractions of dust that is gravitationally bound and weakly scattering.

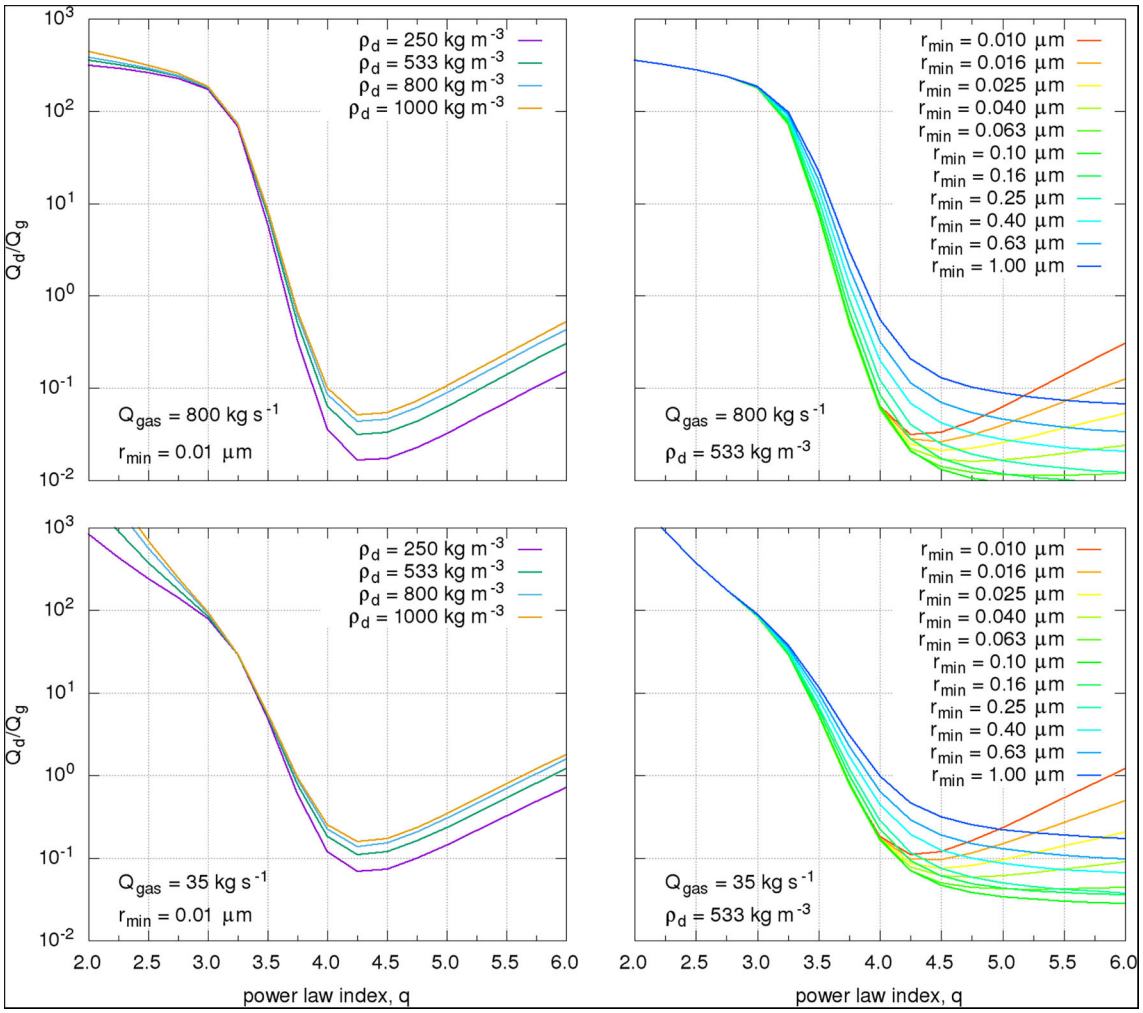


FIGURE 7 | The four panels show the dust-to-gas ration as a function of power-law index. The left two panels show the results for different bulk dust densities (250 kg m^{-3} [purple], 533 kg m^{-3} [green], 800 kg m^{-3} [blue], 1,000 kg m^{-3} [orange]) assuming a minimum dust size of 0.01μm and a gas production rate of 800 kg s^{-1} (top panel, epoch 12) and 35 kg s^{-1} (bottom panel, epoch 6). The two right panels show the results for varying minimum dust size of the power-law from 0.01μm (red) to 1μm (blue) assuming a bulk dust densities of 533 kg m^{-3} and a gas production rate of 800 kg s^{-1} (top panel, epoch 12) and 35 kg s^{-1} (bottom panel, epoch 6).

Two assumptions going into **Figure 6** are worth discussing. First, we have assumed that all dust particles have a bulk density equal to the nucleus density (533 kg m^{-3}). It is likely that the density of small particles is significantly larger than that and that the density then decreases with size. Because the exact relationship of the density as a function of dust size is currently unknown we have not tested a varying density as a function of size. But, we have varied the bulk density for the entire range of dust sizes between 250 kg m^{-3} and 1,000 kg m^{-3}. The two left panels of **Figure 7** show the results for a moderate activity environment (epoch 6—inbound equinox, $Q_g = 35$ kg s^{-1}) and a high activity environment (epoch 12—solstice, $Q_g = 800$ kg s^{-1}). For $3 < q < 3.75$ the differences between the different dust densities is minimal. For $q > 3.75$ the differences are larger, in particular in the high activity case. How the bulk dust density impacts the total dust mass loss will be explored later in this section.

Second, we have currently assumed that the smallest dust size is 0.01μm. This might not be the preferred choice and a much larger smallest size should be considered. The MIDAS instrument detected 1μm particles (e.g., [29]) and there is indirect evidence of sub-micron particles observed by VIRTIS during outbursts [43]. We have thus explored the range of the smallest sizes between 0.01 and 1μm. The two right panels of **Figure 7** explore the effect of the smallest size on the dust-to-gas ratio by varying the smallest size. Compared to the differences seen for different bulk dust density the effect of the smallest size is quite substantial. As we would expect the smallest size does not affect the result for $q < 3$ as in these cases most of the mass is in the large particles. As q increases from 3 the curves for different smallest sizes start diverging. Two trends can be observed. As the smallest

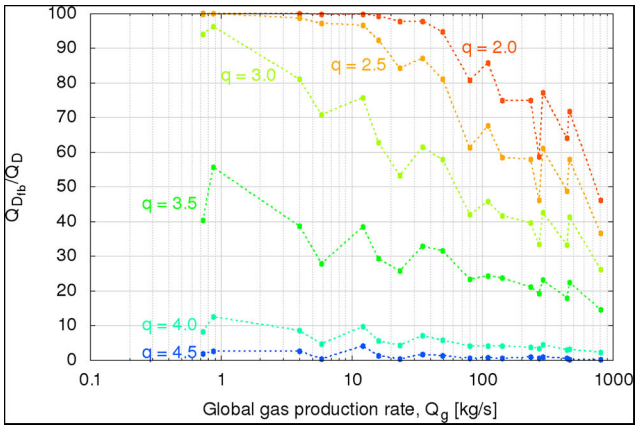


FIGURE 8 | The ratio of dust mass falling back, $Q_{D_{fb}}$ to total dust mass, Q_D is shown as a function of global gas production rate for different power-law exponents.

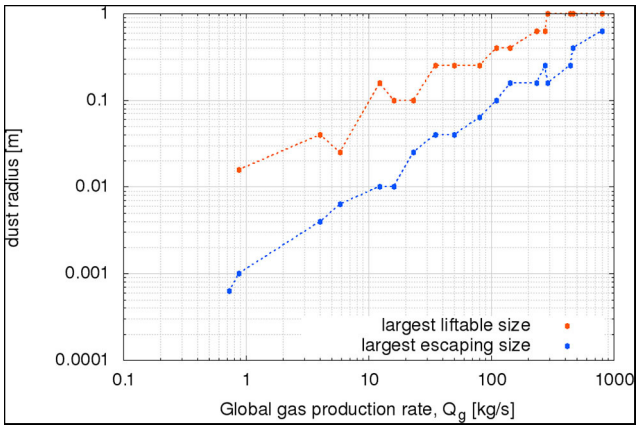


FIGURE 9 | The largest liftable dust size (red curve) and the largest escaping dust size (blue curve) are shown as a function of the gas production rate.

size increases from 0.01μm to $\sim$ 0.1μm the dust-to-gas ratio starts to flatten out beyond $q = 4$. This is caused by the fact that the size distribution is no longer dominated by very inefficiently scattering particles. As the smallest size continues to increase to 1μm the overall dust-to-gas ratio increases. This is because the most efficient scatterers (see **Figure 3**) are being removed from the size distribution and must be compensated by more mass of all other sizes. For very steep power-laws the difference in the dust-to-gas ratio can be up to 1.5 orders of magnitudes. How the smallest size impacts the total dust mass loss will be explored later in this section.

How the fraction of gravitationally bound dust mass, which falls back to the nucleus, varies as a function of global gas production rate is shown in **Figure 8** for different power-law exponents. This illustrates that for very low gas production rates and very shallow power-laws ($q < 2.5$) almost the entire dust mass emitted from the surface will be redeposited. This explains the large increase seen in Q_d/Q_g of **Figure 6**. Conversely, in the case of steep power-laws ($q \geq 4.5$) almost all of the dust escapes the nucleus' gravity field irrespective of the gas production rate. In all cases, the fraction of fall back decreases as the gas production rate increases. Therefore, the fraction of fall back material decreases as the comet approaches the Sun. At large heliocentric distances, large fractions of dust emitted will return to the surface (i.e., $>$ 50% for $q < 3.5$). But the gas and dust production rates are highest at perihelion/solstice, thus the total amount of fall back during one apparition is dominated by the fraction of fall back during that period.

The fraction of fall back is also tightly bound to the maximum liftable dust size. **Figure 9** shows as a function of global gas production rate the largest dust size that can still be lifted from the surface of the nucleus. The figure also shows the largest dust size that can escape the gravity field of the comet. As the gas production rate increases so do the largest liftable and escaping dust sizes. For $Q_g > 300$ kg s^{-1} the largest liftable dust size is larger or equal to 1 m, which is the largest size in our model. Though these sizes, or larger, can be lifted they will not be able to escape the gravity field of the comet and will be redeposited on the surface. The largest size that can escape the comet at peak gas production is roughly 0.6 m. We should note though that this calculation neglects surface cohesion, solar radiation pressure, and heat transport to the subsurface that is needed to eject such large particles. Here we only consider the balance of gas drag and gravity to determine these largest sizes.

The discussion of the previous paragraphs illustrates that multiple properties of the dynamical simulation of the dust coma (size distribution, dust-to-gas ratio, and the fraction of fallback) as well as the optical properties of the dust are not independent but mutually constraining. Example, a given fraction of fallback implies a certain size distribution which in turn constrains the possible dust-to-gas ratios for a particular set of scattering properties. Though for a particular single OSIRIS image these parameters can be constrained there still remains a rather large set of parameters that are consistent with the data as presented to this point (including dust coma brightness in the OSIRIS images and local gas densities of ROSINA/COPS).

While we have only considered constraints within each epoch there is one strong constraint covering the entire mission. That is the measurement of the total mass loss during the Rosetta apparition. During the 2 year mission comet 67P had lost $(10.5 \pm 3.4) \cdot 10^9$ kg [4]. The total mass loss, $M_{tot} = (10.5 \pm 3.4) \cdot 10^9$ kg, is:

$$M_{tot} = M_g + M_d^{esc} \quad , \tag{14}$$

where M_g is the total volatile mass loss, and M_d^{esc} is the total escaping dust mass. For the dust masses, we can further specify that

$$M_d = M_d^{esc} + M_d^{fb} \quad , \tag{15}$$

where M_d is the total dust mass ejected from the nucleus, and M_d^{fb} is the dust mass that falls back to the surface. We have determined the total volatile mass loss to be $(6.1 \pm 1.5) \cdot 10^9$ kg. Combined with the total mass loss of the nucleus it follows that $M_d^{esc} = (4.4^{+4.9}_{-4.4}) \cdot 10^9$ kg. Note that within this interval exists the

possibility that $M_d^{esc} = 0$ kg. Though we know that dust escaped from the nucleus the simple mass balance would not exclude this possibility. We can now integrate the total dust mass loss over the orbit of the comet for different power-law exponents. For the integration, we assume a linear interpolation of the results between epochs. **Figure 10** shows the M_d^{esc} as a function of the power-law exponent, q. Cases, where M_d^{esc} exceeds the nominal dust mass loss of $4.4 \cdot 10^9$ kg (horizontal dashed green line) or the maximum dust mass loss of $9.3 \cdot 10^9$ kg (horizontal dashed red line), can be discarded. Furthermore, where the mass loss curve intersects the mass loss indicates the corresponding power-law that fits the data. **Figure 10** also illustrates the effect of the smallest size—discussed earlier in more detail for an individual OSIRIS image—on the total mass loss. The effect of the smallest size is rather limited for $0.01\mu\text{m} < r_{min} < 1\mu\text{m}$ because the dust mass curves cross between $3.5 < q < 4$. This also implies that the effect of the bulk dust density is even smaller than the effect from the choice of the smallest size (see discussion about **Figure 7** above). We can also see that for $r_{min} >\sim 12\mu$m there will no longer be a nominal solution to the constraints. Further, for $r_{min} >\sim 30\mu$m there is no solution at all because the curve will stay above the maximum escaping mass for all power-law exponents studied here. This means that the minimum dust size must be strictly small than $\sim 30\mu$m and nominally even smaller than $\sim 12\mu$m. **Figure 10** illustrates how we can determine the power-law exponent for the nominal and maximum dust mass loss, which in turn determines the dust-to-gas ratio, dust production rates, fraction of dust fallback. As the minimum size grows larger than 1μm the required power-law exponent increases and becomes rather large. There remains the issue of the minimum escaping mass. As discussed above the lower limit according to the total and volatile mass loss is zero. But for our models, the minimum escaping dust mass is never zero. We have thus chosen the smallest possible mass loss of each model as the minimum mass loss. The resulting power-law exponents, dust mass losses, dust-to-gas ratios, and fall back fraction are summarized in **Table 5**. We have also determined the deposition height, H, that results if the fallback material is spread equally on the smooth deposits (9.43 km^2) identified by Thomas et al. [12].

The results in **Table 5** show that the integrated quantities are rather insensitive to the choice of the smallest dust size if $r_{min} \leq 1\mu$m. For minimum sizes larger than 1μm the power-law becomes steeper and thus the amount of dust fall-back goes down. The dust-to-gas ratio is rather stable for all cases and is of the order of 0.8 with an error of the order of 100%. This means that while the nominal case reflects a comet that contains more volatiles than dust the case of a dusty comet lies within the error.

The fallback in all cases is of the order of 10% and results in a deposition height of the order of 10 cm. Because the deposition is likely non-uniform it is therefore easily thinkable that in certain areas dust of the order of meters is deposited while in others only a few centimeters.

This analysis assumes that the dust size distribution does not change along the orbit. There is an indication (e.g., [44]) that this is not the case and that the slope is varying with heliocentric distance. Our model cannot resolve/constrain this.

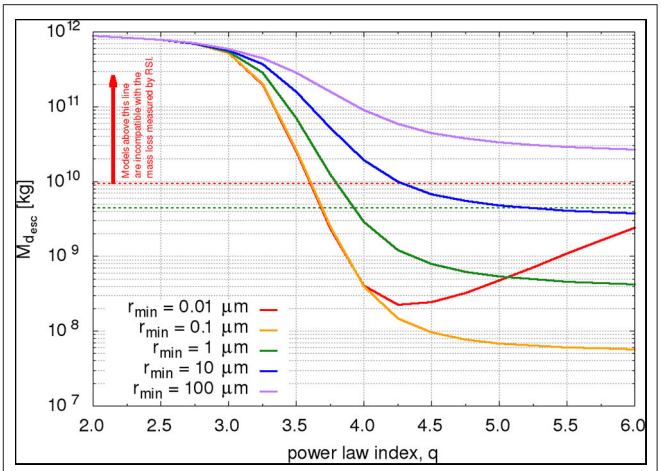


FIGURE 10 | The total escaping dust mass, M_d^{esc}, is shown as a function of the power-law index for five different minimum dust radii, r_{min}. The horizontal dashed lines show the nominal ejected dust mass (green) and the maximum ejected dust mass (red).

TABLE 5 | Power-law exponents, dust mass loss, dust-to-gas ratio, and fall back as a function of the smallest dust size (see **Figure 10**).

r_{min}	q	M_d [kg]	M_d^{esc} [kg]	M_d^{fb} [kg]
0.01μm	$3.7^{+0.57}_{-0.078}$	$5.1^{+6.0}_{-4.9} \cdot 10^9$	$4.4^{+4.9}_{-4.2} \cdot 10^9$	$6.8^{+11}_{-6.8} \cdot 10^8$
0.1μm	$3.7^{+2.3}_{-0.079}$	$5.1^{+6.0}_{-5.0} \cdot 10^9$	$4.4^{+4.9}_{-4.4} \cdot 10^9$	$6.7^{+11}_{-6.7} \cdot 10^8$
1μm	$3.9^{+2.1}_{-0.13}$	$4.7^{+5.6}_{-4.3} \cdot 10^9$	$4.4^{+4.9}_{-4.0} \cdot 10^9$	$3.1^{+7.3}_{-3.1} \cdot 10^8$
10μm	$5.2^{+0.80}_{-0.93}$	$4.5^{+5.1}_{-0.69} \cdot 10^9$	$4.4^{+4.9}_{-0.66} \cdot 10^9$	$0.35^{+1.6}_{-0.35} \cdot 10^8$
100μm	-	-	-	-

r_{min}	M_d/M_g	M_d^{esc}/M_g	M_d^{fb}/M_d [%]	H [cm]
0.01μm	$0.84^{+1.6}_{-0.81}$	$0.73^{+1.3}_{-0.70}$	$13^{+2.6}_{-12}$	14^{+22}_{-14}
0.1μm	$0.84^{+1.6}_{-0.83}$	$0.73^{+1.3}_{-0.72}$	$13^{+2.6}_{-12}$	13^{+21}_{-13}
1μm	$0.78^{+1.5}_{-0.73}$	$0.73^{+1.3}_{-0.67}$	$6.6^{+3.4}_{-5.9}$	$6.3^{+14}_{-6.2}$
10μm	$0.74^{+1.4}_{-0.24}$	$0.73^{+1.3}_{-0.23}$	$0.78^{+1.3}_{-0.78}$	$0.69^{+3.2}_{-0.69}$
100μm	-	-	-	-

Dashed entries mean that no solution is possible for this size.

All the quantities here are heavily dominated by the period around perihelion and summer solstice when the emission was the highest. Therefore, the power indexes found here reflect the values for this period.

The power-law we find to be compatible with the data is an independent result based only on the brightness of the dust coma and the total mass loss balance. Because most mass is ejected around perihelion, this power-law mainly reflects this period and deviations of it at larger heliocentric distances [45] would not influence the result. The value we find is in line with other measurements around perihelion e.g., the *in-situ* measurement of $q = 3.7$ by GIADA [45], $q = 3.8$ by COSIMA [44], as well as ground-based estimates for the dust tail of $3.6 < q < 4.3$ for sizes smaller than 1 mm and $q = 3.6$ for sizes larger than 1 mm [46].

A check of our dust dynamics model is the comparison of our model dust speeds with the ones measured by GIADA. For the

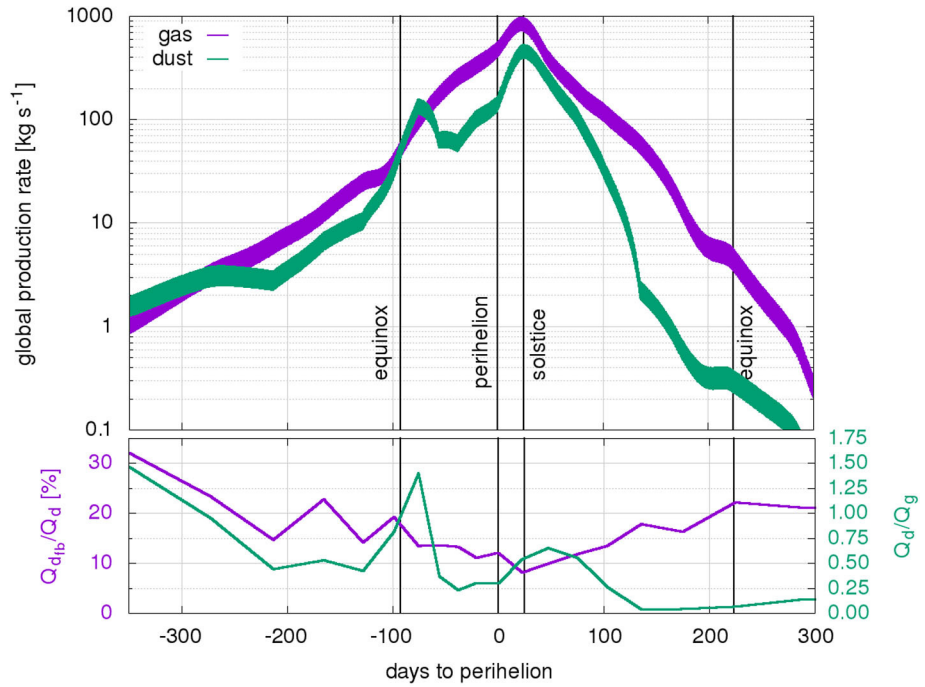


FIGURE 11 | The top panel shows the global production rate for the gas (purple) and dust (green) as a function of days to perihelion. The bands indicate the range due to the diurnal variation. The gas production rates have been constrained by ROSINA/COPS measurements while the dust production rates are from combined constraints of OSIRIS and gas fluxes. For the dust a minimum size of $r_{min} = 0.1$ μm and power-law exponent of $q = 3.75$ are assumed. The bottom panel shows the fraction of dust fall-back (purple) and dust-to-gas ratio (green).

period between 2.2 AU inbound to 2.0 AU outbound [47] have reported 141 dust particle detection for which a dust speed and mass could be determined. Of these, the smallest particle had a mass of $2.8 \cdot 10^{-9}$ kg, which corresponds to a radius of 108 μm assuming a spherical particle with our nominal dust density. The measured dust speeds varied between 0.3 and 34.8 m s^{-1} [47]. A further constraint is the fact that Rosetta spent $\sim$ 65% of the time at phase angles of $\sim$ 90° and an additional $\sim$ 25% of the time at phase angles of $\sim$ 60° which implies that the particles were mainly collected in those locations (see also **Figure 5** in [47]). At a phase angle of 90° our model dust particles with radius 100 μm have speeds of (3.5 ± 1.0) m s^{-1} at the inbound equinox (epoch 6) and (18 ± 5.6) m s^{-1} at the summer solstice (epoch 12). For phase angles of 60° the model dust particles have speeds of (7.0 ± 1.2) m s^{-1} at the inbound equinox and (32 ± 5.2) m s^{-1} at the summer solstice. Our dust speeds are thus well in line with the measured speeds by GIADA given that larger particles will have lower speeds than the ones listed above.

We should highlight that our peak dust production rate ($\sim$ 530 kg s^{-1}) is roughly an order of magnitude lower than those reported by e.g., [46] ($\sim$ 3,000 kg s^{-1}) or [48] ($\sim$ 8,300 kg s^{-1}). Furthermore, [46] report a total dust mass loss of $1.4 \cdot 0^{10}$ kg. As neither [46] nor [48] report any error bars on their results, we cannot asses if they are plausible. If taken at face value both results are inconsistent with the RSI measurement of the total mass loss of the comet [4] taking into account the estimates of the volatile mass loss in this work and others [6, 7].

Finally, the determination of the power-law exponent allows us to determine the dust production rate (**Figure 11**, top panel, green band), dust-to-gas ratio (**Figure 11**, bottom panel, green line), and fraction of fallback (**Figure 11**, bottom panel, purple line) as a function of time. The dust production rates are linearly interpolated between the epochs. Unfortunately, our model is rather noisy but the overall trends are robust enough that we feel comfortable making further conclusions. The fraction of dust fallback is highest at large heliocentric distances and then decreases toward perihelion and reaches its minimum at summer solstice where the activity peaked. Though the faction of fallback is smallest at the peak of the activity (solstice), most mass that is falling back will still be from the period of summer solstice because of the high activity. The behavior of the fraction of dust fallback is symmetric for the inbound and outbound part of the comets' orbit. Contrary to that the dust-to-gas ratio is highest ($\sim$ 1.5) at large heliocentric distances inbound and keeps decreasing along the entire orbit and does not significantly increase on the outbound leg but rather flattens out at $\sim$ 0.1. This might be indicative of the comet shedding its dust mantle, in particular in the northern hemisphere. This trend of decreasing dust-to-gas ratio along the orbit manifests itself also in the asymmetry of the global dust production. To first order, the dust production rate follows the gas production rate during the inbound leg but the dust production rate drops faster than the gas production rate post solstice. This is also observed in ground-based measurements [46]. There is also an intriguing spike in the

dust-to-gas ratio after the inbound equinox coinciding with an increase in the total dust production rate. Future in-depth work will be needed to confirm the nature of this feature which does not seem to be present in the observations of the outer coma from ground-based measurements. But if it is truly there it can be understood as the comet shedding its southern dust mantle because the feature coincides with the period when the southern hemisphere receives increasing insolation.

5. SUMMARY AND CONCLUSIONS

In this work, we have simulated the inner gas and dust coma of comet 67P covering the entire Rosetta mission by splitting it into 20 epochs. The gas production rates of each epoch were constrained by *in-situ* measurements of the gas density by ROSINA/COPS. From that, the total gas mass loss is estimated at $(6.1 \pm 1.5) \cdot 10^9$ kg. This is in line with values published in other works as e.g., $(6.3 \pm 2.0) \cdot 10^9$ kg [7] or $(5.8 \pm 1.8) \cdot 10^9$ kg [6]. It also illustrates that it is not necessary to know the surface-emission distribution well to estimate the total global volatile loss.

By simulating synthetic OSIRIS images of the dust coma we showed how the dynamical and optical properties of the dust can be constrained. In particular, we showed how the dust-to-gas mass production rate ratio, Q_d/Q_g, the power-law exponent, q, the fraction of dust fall back, Q_d^{fb}, and the scattering properties are inter-related and constrain each other. Because these parameters are not independent they need to be fit simultaneously. Example, the lowest mass needed to match the brightness of the dust coma as observed by OSIRIS is achieved with power-law distributions with exponents between 4 and 4.5. Using the constraint of the total mass loss of the comet during the 2015 apparition we were able to show that only a narrow parameter set fits all observations. We determined that power-laws with $q = 3.7^{+0.57}_{-0.078}$ are consistent with the data. This results in a total of $5.1^{+6.0}_{-4.9} \cdot 10^9$ kg of dust being ejected from the nucleus surface, of which $4.4^{+4.9}_{-4.2} \cdot 10^9$ kg escape to space and $6.8^{+11}_{-6.8} \cdot 10^8$ kg (or an equivalent of 14^{+22}_{-14} cm over the smooth regions) is re-deposited on the surface. This leads to a dust-to-gas ratio of $0.73^{+1.3}_{-0.70}$ for the escaping material and $0.84^{+1.6}_{-0.81}$ for the ejected material. Further, the minimum dust size must be strictly smaller than $\sim 30\mu$m and nominally even smaller than $\sim 12\mu$m. We have found that these results are robust with respect to varying the smallest dust size between 0.01 and 1μm and variations in the bulk density of the dust between $250 - -1,000$ kg m^{-3}.

It remains an open question as to how dust particles are lifted/ejected from cometary surfaces [see e.g., 50]. Furthermore, a more detailed study of the change in the dust size distribution with heliocentric distance would be of great interest and could refine the work presented here. Finally, comprehensive work on estimating the amount of dust deposition through e.g., local digital terrain modeling [e.g., method by 49] would provide valuable additional constraints.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author. The data of some figures is available on www.spaceMarschall.net.

AUTHOR CONTRIBUTIONS

RM performed the modeling of the gas and dust coma as well as the comparisons with ROSINA/COPS and OSIRIS and the analysis related to that. JM performed the calculations of the scattering properties of the dust particles. S-BG implemented new features that optimize `DRAG3D`. OP-R produced the unstructured simulation grid within which all `UltraSPARTS` and `DRAG3D` were run. NT wrote the IDL programs for reading and analysis of the OSIRIS images. J-SW provided support for running and optimization of `UltraSPARTS`. All authors contributed to the article and approved the submitted version.

FUNDING

RM acknowledges the support from the Swiss National Science Foundation grant 184482. JM acknowledges the support from ERC Grant No. 757390. The team from the University of Bern is supported through the Swiss National Science Foundation, and through the NCCR PlanetS.

ACKNOWLEDGMENTS

RM acknowledges the support from the Swiss National Science Foundation grant 184482.

JM acknowledges the support from ERC Grant No. 757390.

The team from the University of Bern is supported through the Swiss National Science Foundation, and through the NCCR PlanetS.

We thank Frank Preusker and Frank Scholten for providing us the comet shape model SHAP7 [2] used in this work.

We thank Vladimir Zakharov for providing valuable comments on the section of the analytical solution.

We acknowledge the personnel at ESA's European Space Operations Center (ESOC) in Darmstadt, Germany, European Space Astronomy Center (ESAC) in Spain, and at ESA for the making the Rosetta mission possible. Furthermore, we thank the OSIRS and ROSINA instrument and science teams for their hard work. We thank Martin Rubin and Kathrin Altwegg for giving us access and support to/for the ROSINA/COPS data.

REFERENCES

1. Sierks H, Barbieri C, Lamy PL, Rodrigo R, Koschny D, Rickman H, et al. On the nucleus structure and activity of comet 67P/Churyumov-Gerasimenko. *Science*. (2015) **347**:1044. doi: 10.1126/science.aaa1044
2. Preusker F, Scholten F, Matz KD, Roatsch T, Hviid SF, Mottola S, et al. The global meter-level shape model of comet 67P/Churyumov-Gerasimenko. *Astron Astrophys*. (2017) **601**:L1. doi: 10.1051/0004-6361/201731798
3. Thomas N, Sierks H, Barbieri C, Lamy PL, Rodrigo R, Rickman H, et al. The morphological diversity of comet 67P/Churyumov-Gerasimenko. *Science*. (2015) **347**:440. doi: 10.1126/science.aaa0440

4. Pätzold M, Andert TP, Hahn M, Barriot JP, Asmar SW, Häusler B, et al. The Nucleus of comet 67P/Churyumov-Gerasimenko - Part I: the global view - nucleus mass, mass-loss, porosity, and implications. *Mthly Notices R Astron Soc*. (2019) **483**:2337–46. doi: 10.1093/mnras/sty3171
5. Fougere N, Altwegg K, Berthelier JJ, Bieler A, Bockelée-Morvan D, Calmonte U, et al. Three-dimensional direct simulation Monte-Carlo modeling of the coma of comet 67P/Churyumov-Gerasimenko observed by the VIRTIS and ROSINA instruments on board Rosetta. *Astron Astrophys*. (2016) **588**:A134. doi: 10.1051/0004-6361/201527889
6. Läuter M, Kramer T, Rubin M, Altwegg K. Surface localization of gas sources on comet 67P/Churyumov-Gerasimenko based on DFMS/COPS data. *Mthly Notices R Astron Soc*. (2018) **483**:852–61. doi: 10.1093/mnras/sty3103
7. Combi M, Shou Y, Fougere N, Tenishev V, Altwegg K, Rubin M, et al. The surface distributions of the production of the major volatile species, H2O, CO2, CO and O2, from the nucleus of comet 67P/Churyumov-Gerasimenko throughout the Rosetta Mission as measured by the ROSINA double focusing mass spectrometer. *Icarus*. (2020) **335**:113421. doi: 10.1016/j.icarus.2019.113421
8. Migliorini A, Piccioni G, Capaccioni F, Filacchione G, Bockelée-Morvan D, Erard S, et al. Water and carbon dioxide distribution in the 67P/Churyumov-Gerasimenko coma from VIRTIS-M infrared observations. *Astron Astrophys*. (2016) **589**:A45. doi: 10.1051/0004-6361/201527661
9. Bockelée-Morvan D, Crovisier J, Erard S, Capaccioni F, Leyrat C, Filacchione G, et al. Evolution of CO2, CH4, and OCS abundances relative to H2O in the coma of comet 67P around perihelion from Rosetta/VIRTIS-H observations. *Mthly Notices R Astron Soc*. (2016) **462**(Suppl_1):S170–83. doi: 10.1093/mnras/stw2428
10. Marshall DW, Hartogh P, Rezac L, von Allmen P, Biver N, Bockelée-Morvan D, et al. Spatially resolved evolution of the local H_2O production rates of comet 67P/Churyumov-Gerasimenko from the MIRO instrument on Rosetta. *Astron Astrophys*. (2017) **603**:A87. doi: 10.1051/0004-6361/201730502
11. Biver N, Bockelée-Morvan D, Hofstadter M, Lellouch E, Choukroun M, Gulkis S, et al. Long-term monitoring of the outgassing and composition of comet 67P/Churyumov-Gerasimenko with the Rosetta/MIRO instrument. *Astron. Astrophys*. (2019) **630**:A19. doi: 10.1051/0004-6361/201834960
12. Thomas N, El Maarry MR, Theologou P, Preusker F, Scholten F, Jorda L, et al. Regional unit definition for the nucleus of comet 67P/Churyumov-Gerasimenko on the SHAP7 model. *Planet Space Sci*. (2018) **164**:19–36. doi: 10.1016/j.pss.2018.05.019
13. Thomas N, Davidsson B, El-Maarry MR, Fornasier S, Giacomini L, Gracia-Berná AG, et al. Redistribution of particles across the nucleus of comet 67P/Churyumov-Gerasimenko. *Astron Astrophys*. (2015) **583**:A17. doi: 10.1130/abs/2016AM-281342
14. Keller HU, Barbieri C, Lamy P, Rickman H, Rodrigo R, Wenzel KP, et al. OSIRIS the scientific camera system onboard Rosetta. *Space Sci Rev*. (2007) **128**:433–506. doi: 10.1007/s11214-006-9128-4
15. Balsiger H, Altwegg K, Bochsler P, Eberhardt P, Fischer J, Graf S, et al. Rosina Rosetta orbiter spectrometer for ion and neutral analysis. *Space Sci Rev*. (2007) **128**:745–801. doi: 10.1007/s11214-006-8335-3
16. Marschall R, Su CC, Liao Y, Thomas N, Altwegg K, Sierks H, et al. Modelling observations of the inner gas and dust coma of comet 67P/Churyumov-Gerasimenko using ROSINA/COPS and OSIRIS data: first results. *Astron Astrophys*. (2016) **589**:A90. doi: 10.1051/0004-6361/201628085
17. Marschall R, Mottola S, Su CC, Liao Y, Rubin M, Wu JS, et al. Cliffs versus plains: Can ROSINA/COPS and OSIRIS data of comet 67P /Churyumov-Gerasimenko in autumn 2014 constrain inhomogeneous outgassing? *Astron Astrophys*. (2017) **605**:A112. doi: 10.1051/0004-6361/201730849
18. Gerig SB, Marschall R, Thomas N, Bertini I, Bodewits D, Davidsson B, et al. On deviations from free-radial outflow in the inner coma of comet 67P /Churyumov-Gerasimenko. *Icarus*. (2018) **311**:1–22. doi: 10.1016/j.icarus.2018.03.010
19. Marschall R, Rezac L, Kappel D, Su CC, Gerig SB, Rubin M, et al. A comparison of multiple Rosetta data sets and 3D model calculations of 67P/Churyumov-Gerasimenko coma around equinox (May 2015). *Icarus*. (2019) **328**:104–26. doi: 10.1016/j.icarus.2019.02.008
20. Su CC. *Parallel Direct Simulation Monte Carlo (DSMC) Methods for Modeling Rarefied Gas Dynamics*. Taiwan: National Chiao Tung University (2013).
21. Wu JS, Lian YY. Parallel three-dimensional Direct Simulation Monte Carlo method and its applications. *Comput Fluids*. (2003) **32**:1133–60. doi: 10.1016/S0045-7930(02)00083-X
22. Wu JS, Tseng KC, Wu FY. Parallel three-dimensional DSMC method using mesh refinement and variable time-step scheme. *Comput Phys Commun*. (2004) **162**:166–87. doi: 10.1016/j.cpc.2004.07.004
23. Wu JS, Tseng KC. Parallel DSMC method using dynamic domain decomposition. *Int J Num Methods Eng*. (2005) **63**:37–76. doi: 10.1002/nme.1232
24. Bieler A, Altwegg K, Balsiger H, Berthelier JJ, Calmonte U, Combi M, et al. Comparison of 3D kinetic and hydrodynamic models to ROSINA-COPS measurements of the neutral coma of 67P/Churyumov-Gerasimenko. *Astron Astrophys*. (2015) **583**:A7. doi: 10.1051/0004-6361/201526178
25. Zakharov V, Crifo JF, Rodionov AV, Rubin M, Altwegg K. The near-nucleus gas coma of comet 67P/Churyumov-Gerasimenko prior to the descent of the surface lander PHILAE. *Astron Astrophys*. (2018) **618**:A71. doi: 10.1051/0004-6361/201832883
26. Tzou CY. *Calibrations of ROSINA-COPS and Observations at Comet 67P/Churyumov-Gerasimenko*. Universitat Bern, Bern (2017).
27. Finklenburg S, Thomas N, Su CC, Wu JS. The spatial distribution of water in the inner coma of Comet 9P/Tempel 1: comparison between models and observations. *Icarus*. (2014) **236**:9–23. doi: 10.1016/j.icarus.2014.03.032
28. Marschall R, Liao Y, Thomas N, Wu J. Limitations in the determination of surface emission distributions on comets through modelling of observational data - a case study based on Rosetta observations. *Icarus*. (2019) **346**:113742. doi: 10.1016/j.icarus.2020.113742
29. Mannel T, Bentley MS, Boakes PD, Jeszenszky H, Ehrenfreund P, Engrand C, et al. Dust of comet 67P/Churyumov-Gerasimenko collected by Rosetta/MIDAS: classification and extension to the nanometer scale. *Astron Astrophys*. (2019) **630**:A26. doi: 10.1051/0004-6361/201834851
30. Della Corte V, Rotundi A, Zakharov V, Ivanovski S, Palumbo P, Fulle M, et al. GIADA microbalance measurements on board Rosetta: submicrometer- to micrometer-sized dust particle flux in the coma of comet 67P/Churyumov-Gerasimenko. *Astron Astrophys*. (2019) **630**:A25. doi: 10.1051/0004-6361/201834912
31. Zakharov VV, Ivanovski SL, Crifo JF, Della Corte V, Rotundi A, Fulle M. Asymptotics for spherical particle motion in a spherically expanding flow. *Icarus*. (2018) **312**:121–127. doi: 10.1016/j.icarus.2018.04.030
32. Muinonen K, Markkanen J, Väisänen T, Peltoniemi J, Penttilä A. Multiple scattering of light in discrete random media using incoherent interactions. *Optics Lett*. (2018) **43**:683. doi: 10.1364/OL.43.000683
33. Markkanen J, Väisänen T, Penttilä A, Muinonen K. Scattering and absorption in dense discrete random media of irregular particles. *Optics Lett*. (2018) **43**:2925. doi: 10.1364/OL.43.002925
34. Markkanen J, Agarwal J, Väisänen T, Penttilä A, Muinonen K. Interpretation of the phase functions measured by the OSIRIS instrument for comet 67P/Churyumov-Gerasimenko. *Astrophys J Lett*. (2018) **868**:L16. doi: 10.3847/2041-8213/aaee10
35. Markkanen J, Agarwal J. Scattering, absorption and thermal emission by large cometary dust particles: synoptic numerical solution. *Astron Astrophys*. (2019) **631**:A164. doi: 10.1051/0004-6361/201936235
36. Dorschner J, Begemann B, Henning T, Jaeger C, Mutschke H. Steps toward interstellar silicate mineralogy. II. Study of Mg-Fe-silicate glasses of variable composition. *Astron Astrophys*. (1995) **300**:503.
37. Jäger C, Mutschke H, Henning T. Optical properties of carbonaceous dust analogues. *Astron Astrophys*. (1998) **332**:291–9.
38. Bertini I, La Forgia F, Tubiana C, Güttler C, Fulle M, Moreno F, et al. The scattering phase function of comet 67P/Churyumov-Gerasimenko coma as seen from the Rosetta/OSIRIS instrument. *Mthly Notices R Astron Soc*. (2017) **469**:S404–15. doi: 10.1093/mnras/stx1850
39. Fornasier S, Hasselmann PH, Barucci MA, Feller C, Besse S, Leyrat C, et al. Spectrophotometric properties of the nucleus of comet 67P/Churyumov-Gerasimenko from the OSIRIS instrument onboard the ROSETTA spacecraft. *Astron Astrophys*. (2015) **583**:A30. doi: 10.1051/0004-6361/201525901

40. Feller C, Fornasier S, Hasselmann PH, Barucci A, Preusker F, Scholten F, et al. Decimetre-scaled spectrophotometric properties of the nucleus of comet 67P/Churyumov-Gerasimenko from OSIRIS observations. *Mthly Notices R Astron Soc*. (2016) **462**:S287–303. doi: 10.1093/mnras/stw2511
41. Masoumzadeh N, Oklay N, Kolokolova L, Sierks H, Fornasier S, Barucci MA, et al. Opposition effect on comet 67P/Churyumov-Gerasimenko using Rosetta-OSIRIS images. *Astron Astrophys*. (2017) **599**:A11. doi: 10.1051/0004-6361/201629734
42. Laor A, Draine BT. Spectroscopic constraints on the properties of dust in active galactic nuclei. *Astrophys J*. (1993) **402**:441. doi: 10.1086/172149
43. Bockelée-Morvan D, Rinaldi G, Erard S, Leyrat C, Capaccioni F, Drossart P, et al. Comet 67P outbursts and quiescent coma at 1.3 au from the Sun: dust properties from Rosetta/VIRTIS-H observations. *Mthly Notices R Astron Soc*. (2017) **469**(Suppl_2):S443–58. doi: 10.1093/mnras/stx1950
44. Merouane S, Stenzel O, Hilchenbach M, Schulz R, Altobelli N, Fischer H, et al. Evolution of the physical properties of dust and cometary dust activity from 67P/Churyumov-Gerasimenko measured *in situ* by Rosetta/COSIMA. *Mthly Notices R Astron Soc*. (2017) **469**(Suppl_2):S459–74. doi: 10.1093/mnras/stx2018
45. Fulle M, Marzari F, Della Corte V, Fornasier S, Sierks H, Rotundi A, et al. Evolution of the dust size distribution of comet 67P/Churyumov-Gerasimenko from 2.2 au to Perihelion. *Astrophys J*. (2016) **821**:19. doi: 10.3847/0004-637X/821/1/19
46. Moreno F, Muñoz O, Gutiérrez PJ, Lara LM, Snodgrass C, Lin ZY, et al. The dust environment of comet 67P/Churyumov-Gerasimenko: results from Monte Carlo dust tail modelling applied to a large ground-based observation data set. *Mthly Notices R Astron Soc*. (2017) **469**(Suppl_2):S186–94. doi: 10.1093/mnras/stx1424
47. Della Corte V, Rotundi A, Fulle M, Ivanovski S, Green SF, Rietmeijer FJM, et al. (2016) 67P/C-G inner coma dust properties from 2.2 au inbound to 2.0 au outbound to the Sun. *Mthly Notices R Astron Soc*. 462:S210–9. doi: 10.1093/mnras/stw2529
48. Ott T, Drolshagen E, Koschny D, Güttler C, Tubiana C, Frattin E, et al. Dust mass distribution around comet 67P/Churyumov-Gerasimenko determined via parallax measurements using Rosetta's OSIRIS cameras. *Mthly Notices R Astron Soc*. (2017) **469**:S276–84. doi: 10.1093/mnras/stx1419
49. Jorda L, Gaskell R, Capanna C, Hviid S, Lamy P, Durech J, et al. The global shape, density and rotation of Comet 67P/Churyumov-Gerasimenko from preperihelion Rosetta/OSIRIS observations. *Icarus*. (2016) **277**:257–78. doi: 10.1016/j.icarus.2016.05.002
50. Vincent, J.-B., Farnham T, Kuhrt E, Skorov Y, Marschall R, Oklay N, et al. Local manifestations of cometary activity. *Space Sci Rev*. (2019) **215**:30. doi: 10.1007/s11214-019-0596-8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership