



Interlaboratory comparison of testing hydraulic, elastic, and failure properties in compression: lessons learned

Yan Cheng¹ · David Lockner² · Mandy Duda¹ · Carolyn Morrow² · Demian Saffer^{3,4} · Insun Song^{3,5} · Jörg Renner¹

Received: 15 August 2023 / Accepted: 2 September 2023 / Published online: 4 October 2023
© The Author(s) 2023

Abstract

Many geoscientific problems require us to exploit synergies of experimental and numerical approaches, which in turn lead to questions regarding the significance of experimental details for validation of numerical codes. We report results of an interlaboratory comparison regarding experimental determination of mechanical and hydraulic properties of samples from five rock types, three sandstone varieties with porosities ranging from 5% to 20%, a marble, and a granite. The objective of this study was to build confidence in the participating laboratories' testing approaches and to establish tractable standards for several physical properties of rocks. We addressed the issue of sample-to-sample variability by investigating the variability of basic physical properties of samples of a particular rock type and by performing repeat tests. Compressive strength of the different rock types spans an order of magnitude and shows close agreement between the laboratories. However, differences among stress–strain relations indicate that the external measurement of axial displacement and the determination of system stiffness require special attention, apparently more so than the external load measurement. Furthermore, post-failure behavior seems to exhibit some machine-dependence. The different methods used for the determination of hydraulic permeability, covering six orders of magnitude for the sample suite, yield differences in absolute values and pressure dependence for some rocks but not for others. The origin of the differences in permeability, in no case exceeding an order of magnitude, correlate with the compressive strength and potentially reflect a convolution of end plug–sample interaction, sample-to-sample variability, heterogeneity on sample scale, and/or anisotropy, the last two aspects are notably not accounted for by the applied evaluation procedures. Our study provides an extensive data set apt for “benchmarking” considerations, be it regarding new laboratory equipment or numerical modeling approaches.

Keywords Static Young's modulus · Rock strength · Permeability

Introduction

Results from laboratory tests on rock samples are critical for the derivation and substantiation of constitutive models to be used in modeling beyond the spatial and temporal scales of laboratory and field tests (Kolditz et al. 2021). The synergies between experimental and numerical approaches (e.g., Esterhuizen 2014) range from hazard prevention, in the context of volcano activity (Heap and Violay 2021), rockbursts (Li et al. 2019; Wang et al. 2021), and waste repositories (e.g., Bossart 2007), to initiatives to build virtual rock physics laboratories for educational purposes (Zhu et al. 2012; Vanorio et al. 2014). The comparability of results obtained in standardized experiments forms the basis for the credibility of laboratory work. The demands on the experimental procedure are particularly high in geosciences and geotechnical

✉ Yan Cheng
yan.cheng@rub.de

¹ Institute for Geology, Mineralogy, and Geophysics, Ruhr-Universität Bochum, Bochum, Germany

² U.S. Geological Survey, Menlo Park, USA

³ Department of Geosciences, Pennsylvania State University, State College, USA

⁴ Present Address: Department of Geological Sciences, Jackson School of Geosciences, Austin, TX, USA

⁵ Present Address: Korea Institute of Geoscience & Mineral Resources, 92 Gwahang-No, Yuseong, Daejeon 305-350, Korea

engineering, because the investigated rock material is often heterogeneous, anisotropic, and limited in its quantity.

In economic applications, subsurface characterization rests on standardized preliminary surveys to plan processes and costs based on results gained under comparable conditions. Examples of regulations serving this purpose are standards published by the American Society of Testing Materials (e.g., ASTM 2017), suggested methods published by the International Society of Rock Mechanics (ISRM; e.g., Kovari et al. 1983), or national standards and recommendations. In scientific context, the investigated problems are usually highly specialized and require deviations from such standards. Intermediate and deep (core) drilling operations certainly represent an endmember among geoscientific projects, because costs are extremely high and the resulting sample material is severely limited. Such drilling operations became increasingly important during the past decades, for example, regarding nuclear waste disposal (e.g., Almén 1994; Delay et al. 2007), mitigation of geohazard (e.g., Prior and Doyle 1984) or geothermal energy provision (e.g., Fridleifsson and Elders 2005). These endeavors may benefit from a process understanding that cannot be gained from material and structure characterization based on field surveys and laboratory tests alone, but require a combination of field testing and large-scale modeling. The complexity of the modeling, both in terms of structures and relevant processes, often mandates the use of numerical codes that have to be verified, validated, and benchmarked using independent constraints from experiments and observational evidence spanning scales from hand samples to rock masses (e.g., Jing 2003; Diehl et al. 2019; Birkholzer and Bond 2022).

It is not uncommon that individual studies combine dedicated experimental work and numerical modeling of rock failure behavior in general (e.g., Holt et al. 2005) or during engineering operations, such as hydraulic fracturing (Deb et al. 2021) and tunneling (Zhang et al. 2018). To tap the large pool of the results of independent experimental studies, a rigorous assessment of the significances of their outcomes may lead to improved understanding of fundamental questions related to the role of methodological peculiarities vs. that of sample-to-sample variability. Comparative studies differ regarding the number of involved laboratories, considered rock varieties, and applied methods (Appendix A), with a good fraction dedicated to the specific and difficult task of determining hydraulic properties of close to impermeable shales (e.g., Ghanizadeh et al. 2015), for which a qualitative method comparison is provided by Sander et al. (2017). Often, different methods for determination of a particular property are compared by tests in a single laboratory on a single sample, at times even in a single device (e.g., Winhausen et al. 2021; Schepp and Renner 2021; Zhang et al. 2022). Efforts regarding interlaboratory validation tests

are documented from the 1980ies, but partly in reports to funding agencies (e.g., Rasilainen et al. 1996; Sandström 2006) or in conference papers (e.g., McPhee and Arthur 1994; Davy et al. 2019) causing problems to track details. True round robins, in principle possible for non-destructive testing (e.g., Rasilainen et al. 1996; Profice et al. 2016), eliminate sample-to-sample variability and thus allow for assessing the role of protocol deviations and method principles, but pose organizational challenges and raise questions regarding history dependence of measurement results. These challenges are probably the reasons for the up to today largest comparative study involving 24 laboratories refraining from attempting a round robin for hydraulic permeability testing of Grimsel granodiorite (David et al. 2018a,b). For destructive strength testing (e.g., Pincus 1993, 1994, 1996; Minardi et al. 2021), however, one has to resort to the selection of to-be-distributed sample suites based on their a-priori characterization (e.g., Minardi et al. 2021), accompanied by the challenge to minimize the uncertainty of the role of sample-to-sample variability, for example, by centralized sample preparation and characterization. In cases, previous studies tended to focus on statistical analyses of results omitting a rigorous uncertainty analysis of the individual measurements (e.g., David et al. 2018a, 2018b), hampering the assessment of the significance of observed differences. For the present study, rock mechanics and rock physics laboratories worldwide were invited to participate in an interlaboratory comparison in the context of the San Andreas Fault Observatory at Depth (SAFOD) deep drilling project (Lockner et al. 2009; Logan et al. 2010; Zoback et al. 2010). Test conditions and aspects of procedures were specified before laboratories received sample blocks from five different rock types. The five rock types were selected, because they (i) occur in deposits with sizes motivating commercial quarrying and thus promise future availability, (ii) have been subject of a range of previous studies and accordingly were expected to span a wide range in the physical properties to be investigated, and (iii) promised to minimize the influence of anisotropy and to ensure homogeneity at the decimeter-scale to allow for preparation of comparable samples. Owing to the destructive nature of strength tests and potential irreversible interactions between fluid and samples, we refrained from a round robin procedure, but the group at the U.S. Geological Survey, Menlo Park, (USGS) organized selection, purchase, and shipment of blocks of the rock types, from which the participating institutions prepared samples locally. The specific objectives of this study were.

- (1) To compare the experimental approaches—including sample preparation—and results from different laboratories to determine causes for potential deviations among results,

- (2) To establish tractable standards required for research objectives associated with deep drilling projects,
- (3) To establish the significance of results of laboratory tests in the light of verification and validation efforts for numerical models, and
- (4) To build confidence in the laboratories' procedures.

We provide results for Young's modulus and compressive strength derived from uniaxial and triaxial deformation experiments of intact rock samples (U.S. Geological Survey—USGS, Ruhr-Universität Bochum—RUB) and for hydraulic permeability (USGS, RUB, The Pennsylvania State University—PSU), the central physical properties for hydromechanical modeling whose importance for

fundamental research and industrial applications is increasingly appreciated (e.g., Neuzil 2003; Ghassemi 2012).

Materials and methods

Materials

Five rock types were studied (Fig. 1, Table 1):

- (1) Crab Orchard sandstone is a fine-grained, low-porosity sandstone mainly composed of quartz (91.0 wt%), minor amounts of plagioclase (0.5 wt%) and orthoclase (1.5 wt%), and clay minerals (smectite: 3.0 wt%, illite/

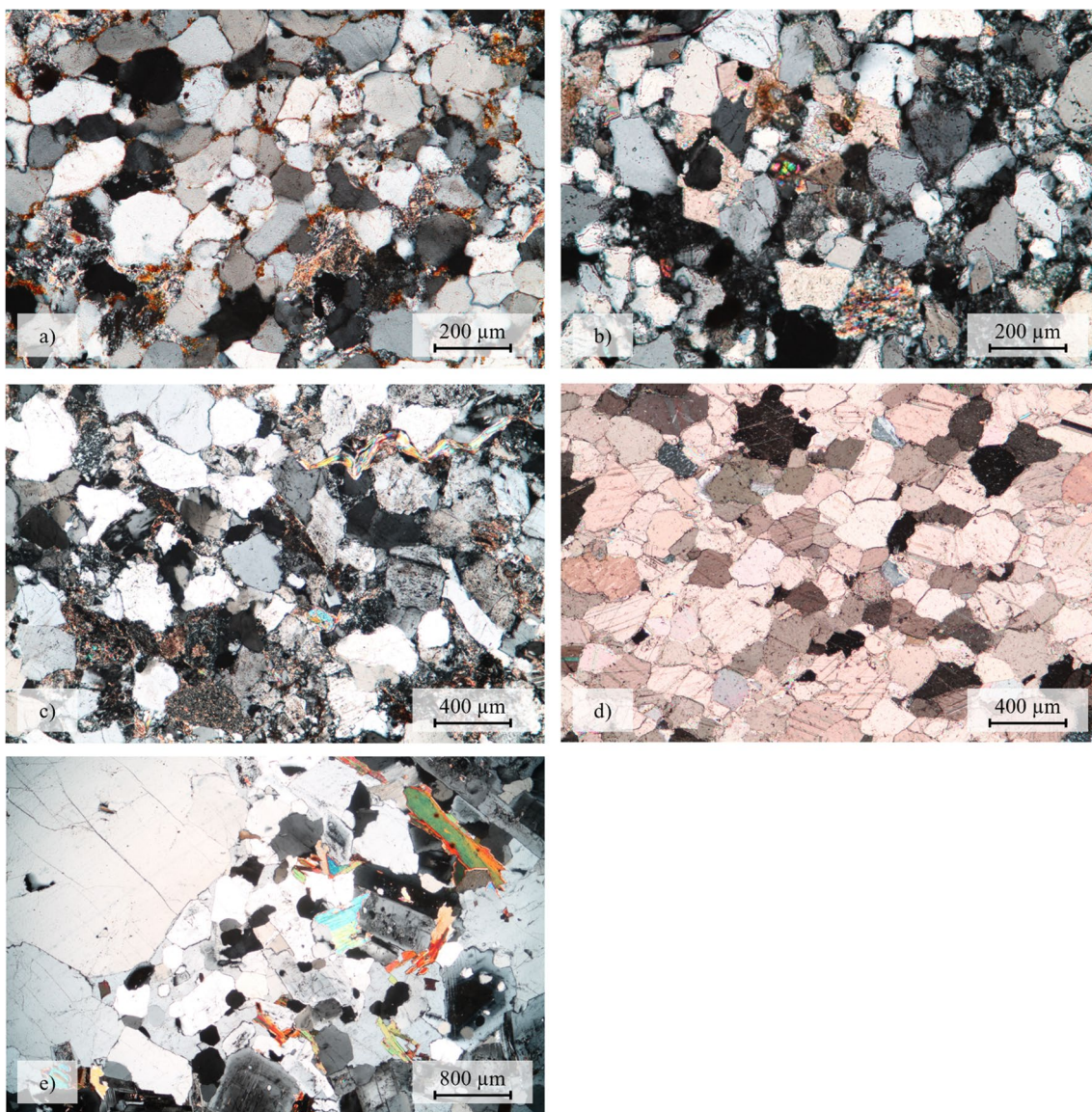


Fig. 1 Optical micrograph images (crossed polarized light) of **a** Crab Orchard sandstone, **b** Berea sandstone, **c** Wilkeson sandstone, **d** Carrara marble, and **e** Sierra White granite. Images are taken from thin sections prepared perpendicular to the drilling directions for the samples

Table 1 Mineralogical compositions from X-ray diffraction in weight percentage (wt%) with uncertainties of ± 2 –5% depending on the mineral

Mineral	Crab Orchard	Berea	Wilkeson	Carrara	Sierra White
Quartz	91.0	88.0	50.0	Traces	38.0
Orthoclase	1.5	5.0	10.0		10.0
Plagioclase	0.5	2.0	26.0		43.0
Calcite				100.0	
Dolomite		2.5	4.0		
Siderite			2.0		
Micas			8.0		9.0
Illite/Muscovite	4.0				
Smectite	3.0				
Kaolinite		2.5			

muscovite: 4.0 wt%). The nominal porosity is about 5% (Benson et al. 2005) and grain size as inferred from thin sections is $< 300 \mu\text{m}$ with an average and a standard deviation of $79 \pm 11 \mu\text{m}$ (line-intercept).

- (2) Berea sandstone represents a lightly banded sandstone with about 20% nominal porosity (e.g., Churcher et al. 1991) composed of 88.0 wt% quartz, 5.0 wt% orthoclase, 2.0 wt% plagioclase, 2.5 wt% dolomite, and 2.5 wt% kaolinite. The grain size distribution is similar to that of Crab Orchard sandstone with an average and standard deviation of $98 \pm 12 \mu\text{m}$ and quartz grains not exceeding $300 \mu\text{m}$.
- (3) Wilkeson sandstone represents a medium-grained sandstone with 10% nominal porosity (e.g., Duda and Renner 2013). It is composed of 50 wt% quartz, 10 wt% orthoclase, 26 wt% plagioclase, 4 wt% dolomite, 2 wt% siderite, and 8 wt% mica. The maximum grain size of individual quartz grains reaches up to 2 mm with an average and standard deviation of $172 \pm 27 \mu\text{m}$.
- (4) Carrara marble is composed of medium-grained calcite with minor additional constituents and a low porosity (e.g., Schmid et al. 1980). A minor grain size anisotropy barely exceeding the standard deviation was determined with $150 \pm 20 \mu\text{m}$ and $173 \pm 22 \mu\text{m}$ in two orthogonal directions.
- (5) Sierra White granite (Knowles granodiorite) is a low-porosity granite (e.g., Miller and Florence 1991) composed of quartz (38 wt%), orthoclase (10 wt%), plagioclase (43 wt%), and micas (9 wt%), including muscovite and biotite. It exhibits a wide range of grain sizes from a few tens of μm to several mm with an average and standard deviation of $649 \pm 257 \mu\text{m}$.

Apart from color gradients for Berea sandstone the investigated blocks showed no macroscopic signs of heavy weathering, anisotropy or heterogeneities.

Sample preparation

Uniaxial and triaxial deformation tests and permeability tests were performed on cylindrical samples prepared by the individual groups who were provided with blocks of the various rock types, whose faces were labelled by the group at USGS as T–B, N–S, and E–W. Specimens were drilled with water-cooled diamond drill bits. All samples intended for comparative measurements were cored in the T–B orientation uniformly defined for all participating institutions, but samples of Wilkeson sandstone for permeability measurements at PSU that were drilled in E–W direction, i.e., orthogonal to the “standard direction”. At PSU, additional samples for permeability measurements were drilled from Berea sandstone and Crab Orchard sandstone in E–W and N–S directions.

For strength tests, right cylinders were prepared (USGS: 25.4 mm diameter \times 63.5 mm length and RUB: 30 mm diameter \times 75 mm length), providing an aspect ratio of about 2.5:1, chosen to ensure a homogeneous stress distribution in the center of samples when subjected to conventional compression (Paterson and Wong 2005). Samples for permeability tests had nominal dimensions of 25.4 mm diameter \times 50 mm length (USGS, PSU) and 30 mm diameter \times 50 mm length (RUB). For both tests, end faces were ground square to within 0.1% parallelism. At the USGS, samples were additionally cylindrically ground to achieve a uniform diameter (within $\pm 0.01 \text{ mm}$) and consistent surface finish, after which they were cleaned with acetone. Samples prepared at RUB by drilling only exhibited diameter variations of less than $\pm 0.03 \text{ mm}$ and were devoid of drilling-score marks. Diameters were measured by calipers with a resolution and an accuracy of better than 0.01 and 0.1 mm, respectively. Finished samples were vacuum-dried at $\sim 60 \text{ }^\circ\text{C}$ for approximately 24 h.

Except for Sierra White granite, the diameter of the specimens exceeded the largest grains in the rock by at least a factor of six, in agreement with ISRM’s suggested methods (Bieniawski and Bernede 1979). In the light of this favorable size vs. grain size ratio, deviations of sample size from recommendations for deformation tests (e.g., ASTM 2017) were allowed on purpose—all samples were smaller than the recommended 40 to 50 mm in diameter—to account for requirements of testing apparatus and to simulate typical material limitations associated with scientific drilling projects.

Sample-to-sample variability deduced from basic rock physical properties

Prepared samples were investigated for their basic physical properties at RUB to exemplarily assess sample-to-sample variability. The differences in basic physical properties of samples originating from a specific block and determined at ambient conditions were not significant, as standard deviations were generally smaller than the experimental uncertainty determined by error propagation (Table 2). Thus, the five rock types were considered sufficiently homogeneous for the planned experiment series and the comparison among laboratories.

Experimental procedures

All tests were to be performed according to instructions concerning sample treatment, number of repeat tests, and applied pressures and their sequences (Tables 3, 4). We refer the reader to Lockner (1998), Duda and Renner (2013), and Ahrens et al. (2017) for technical details of the apparatuses used for deformation tests. The testing procedures did not fully comply with ISRM’s suggested methods (Kovari et al. 1983): (a) spherical seats were not employed; (b) tests were run in displacement control selecting piston velocities in the two laboratories that resulted in the pre-described strain rate of $\sim 1 \times 10^{-5} \text{ s}^{-1}$ for the samples with

Table 2 Average values (avg), standard deviations (std), and experimental uncertainty (Δ) of density ρ , P-wave and S-wave velocities (v_p and v_s) of dry and saturated (sat) samples, and connected porosity ϕ for each rock type. The number of investigated samples is indicated in parenthesis

		Crab Orchard		Carrara		Berea		Wilkeson		Sierra White	
		avg \pm std	Δ	avg \pm std	Δ	avg \pm std	Δ	avg \pm std	Δ	avg \pm std	Δ
ρ	(kg/m ³)	2536 \pm 3 (13)	10	2709 \pm 2 (9)	11	2161 \pm 3 (13)	9	2452 \pm 2 (9)	10	2642 \pm 2 (12)	11
$v_{p,dry}$	(m/s)	3645 \pm 77 (13)	151	4954 \pm 234 (9)	241	2558 \pm 27 (13)	74	3013 \pm 15 (9)	105	4165 \pm 73 (12)	192
$v_{s,dry}$	(m/s)	2608 \pm 54 (8)	195	3007 \pm 94 (8)	259	–	–	1889 \pm 75 (9)	122	2553 \pm 105 (12)	193
$v_{p,sat}$	(m/s)	4492 \pm 95 (4)	297	–	–	3334 \pm 15 (5)	153	3799 \pm 6 (3)	217	5598 \pm 50 (3)	460
$v_{s,sat}$	(m/s)	3826 \pm 151 (4)	631	–	–	2543 \pm 99 (4)	279	2854 \pm 143 (3)	357	4181 \pm 78 (3)	750
ϕ	(%)	5.3 \pm 0.3 (4)	0.4	–	–	18.3 \pm 0.2 (13)	0.2	8.7 \pm 0.1 (9)	0.3	1.2 \pm 0 (3)	0.4

Table 3 Specifications for deformation experiments (after preparation)

Crab Orchard sandstone and Carrara marble	Sierra White granite	Berea and Wilkeson sandstone
Vacuum-dry samples for 24 h at 60 °C (or 48 h at room temperature)		
Run all tests room-dry		Saturate with distilled water by introducing distilled water into sample while still in vacuum chamber
		Run all tests at constant pore pressure of 2 MPa Record volume change of the sample during deformation
Perform constant strain rate tests at 10^{-5} s^{-1} axial shortening rate		
Confining pressures: 0, 10, 20, 50, 150 MPa	0, 20, 60 MPa*	12, 22, 52, 152 MPa
Repeat tests at confining pressures: 3 at 100 MPa	5 at 40 MPa*	3 at 102 MPa
Determine peak and (if possible) residual strength		

* USGS followed the procedure for Crab Orchard sandstone omitting a uniaxial test

Table 4 Specifications for permeability tests (after preparation)

Sierra White granite, Crab Orchard, Berea and Wilkeson sandstone
Vacuum-dry samples for 24 h at 60 °C (or 48 h at room temperature)
Saturate with distilled water by introducing distilled water into sample while still in vacuum chamber
Run all tests at nominal pore pressure of 10 MPa
After assembling: evacuate to remove stray air bubbles and fill pore pressure system with water
Measure permeability in the sequence of confining pressures at nominal pore pressure, no axial differential stress
Sequence of effective confining pressure, here addressing the unweighted difference between confining and pore pressure: 10, 30, 60, 100, 30, 60, 100, 30, 60, 100 MPa*
Wait 1 h at each effective pressure before measurement to allow sample to relax
Care should be taken to avoid over-pressuring a sample prior to permeability test to avoid complications from hysteresis of permeability
Determine permeability (and, if applicable, storage capacity and hydraulic diffusivity)

*PSU used a different sequence of effective confining pressures: 5, 10, 30, 60, 30, 60, 30, 60 MPa

different lengths (Table 3), and with controlled confining pressure. The true strain rates vary over the course of a test by up to a factor of about 2 between the phases of initial steep stress increase and the near constant stress conditions at maximum stress in a single test and also between stiffest (Carrara marble) and most compliant (Berea and Wilkeson sandstone) samples owing to the system deformation (please see the data availability statement for links to test records).

Four methods were used at the participating institutions to obtain permeability: constant-flow, constant-head, and pulse tests at PSU, constant-head tests at USGS, and oscillatory pore-pressure tests at RUB. For theoretical background and experimental setup of permeability tests, we refer to Bernabé et al. (2006), Song, and Renner (2007), Song et al. (2013), and David et al. (2018a, b).

The necessary steps for the evaluation of the mechanical and hydraulic tests are detailed in Appendix B, including a comprehensive discussion of involved uncertainties. Specifically, the conversions of recorded displacements to strains and recorded loads to stresses and stress differences, the difference between axial stress and confining pressure also referred to as deviatoric or differential stress (see Paterson and Wong 2005), need to account for the (current) sample dimensions and system stiffness. The compliances of the assemblies used at USGS and RUB are about 0.002 mm/MPa and 0.001 mm/MPa, respectively, and thus the corrections involved in strain determination amount to up to 70% of the total recorded displacement for tests at USGS on the stiffest rock type, Carrara marble. The different applied hydraulic methods essentially rest on fitting analytical functions to observed pressure transients or spectral analyses of the periodic pressure signals.

Uncertainty analysis

The principles of the estimation of uncertainties of reported quantities relying on Gaussian error propagation of the accuracies of sensors and parameters are documented in Appendix B. Commercial sensors in the United States (US) are traceable back to National Institute of Standards and Testing (NIST). The European providers of the sensors used at RUB guarantee conformity with DIN EN ISO/IEC 17025 (ISO/IEC 2017), i.e., the regulation for calibration services. We used the sensitivities provided by suppliers when transforming electrical signals to physical quantities. Furthermore, displacement transducers are calibrated on a regular basis against calipers; pressure gauges are referenced to analog Heise gauges; the readings of load cells are checked in relation to pressures recorded during hydrostatic loading of the triaxial rigs, measurements that also constrain the friction on the loading piston, and, at RUB, are also tested against a force ring.

Electronic noise in the digitized signal is small compared to the uncertainty of stress difference as determined by the error analysis. The uncertainty of stress difference of 0.4% calculated for peak and residual strengths (indicated for RUB data in the corresponding figures) includes accuracy of the external load cell and the uncertainty in initial sample diameter, i.e., the uncertainty related to the accuracy of the used caliper and shape imperfections but not the change in cross section due to pressurization or axial shortening. Using only initial cross section ensures the direct comparability of the results from the two laboratories, but leads to an increasing overestimation of stress difference with increasing axial strain (see Appendix B).

Stress difference is calculated relative to axial stress on the moving piston before it contacts the specimen (hit-point);

this procedure eliminates seal friction as a source of uncertainty in axial load but for its potential variability with piston deformation. Yet, results of calibration experiments at hydrostatic conditions and deviatoric loading suggest that the friction on the deformation piston is controlled by the confining pressure and does not change with increasing axial load. Nevertheless, friction on the loading piston constitutes an example of methodological uncertainties that are difficult to constrain precisely and that are also encountered for the other physical property determinations (for details see Appendix B). For permeability determination, a likewise critical methodological issue is, for example, to what extent the combination of sample length and used end-plugs actually approximate the condition of one-dimensional flow underlying the evaluation of pressure transients. We propose that an accuracy in permeability of half an order of magnitude appears a realistic, in cases possibly conservative, rule of thumb. Smaller uncertainties have been reported for permeability (e.g., Benson et al. 2005; David et al. 2018a), but it seems that the full cumulative effect of the various sources of uncertainty was not appreciated in these cases. The partial consideration of uncertainty is potentially acceptable when the objective is to resolve the effect of a specific parameter, such as pressure on permeability, in a single study but not for an interlaboratory comparison.

Results

Mechanical parameters

Apparent Young's modulus

The recorded stress–strain curves exhibit various degrees of non-linearity complicating determination of Young's moduli (Fig. 2). The values reported here, labeled “apparent” to indicate that they might differ from intrinsic Young's moduli, represent the maximum slope of the tangent to a polynomial fit to the pre-peak stress–strain curve. For about half of the tests, the apparent moduli determined by the two institutions agree within 15% (Fig. 2). However, the moduli determined at RUB tend to be larger than the ones determined at USGS. We do not find systematics in the dependencies of the moduli on confining pressure of the deformation tests; for example, the moduli measured at RUB for Carrara marble and Wilkeson sandstone exhibit much less and more pronounced pressure dependence than the ones determined at USGS. Neither do we observe a clear trend in the discrepancies between the moduli from the two laboratories with their absolute values nor between tests on dry and saturated samples, requiring different assemblies.

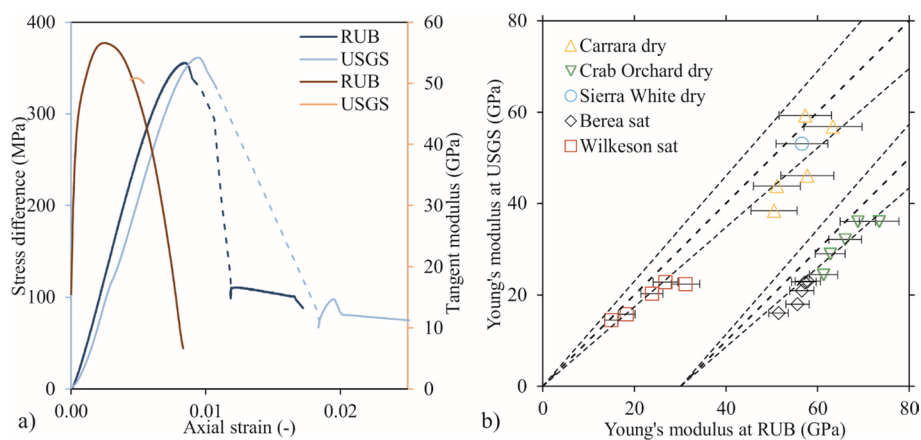


Fig. 2 **a** Examples of stress–strain curves (blue colors) for dry samples of Sierra White granite deformed at 20 MPa confining pressure; the tangent moduli (brown colors) are gained from polyfits to the stress–strain curves (USGS: degree 5, RUB: degree 10), their maxima are used as static Young's moduli. The dashed sections with markers represent the phases of rapid failure, during which the elastic energy stored in the pistons included between the measuring points of the external displacement transducers unloads into the weakening sample. **b** Comparison between maximum tangent modulus, here

used as constraint on apparent (static) Young's modulus, determined by RUB and USGS. Error bars indicate experimental uncertainty. The long-dashed lines indicate one-to-one correspondence and the short-dashed lines indicate 15% boundaries. The data for Berea sandstone and Crab Orchard sandstone, and the corresponding correlation lines are shifted along the x -axis for presentational purposes. (acronyms USGS and RUB denote data gained at U.S. Geological Survey, Menlo Park, and Ruhr-Universität Bochum, respectively; labels “dry” and “sat” distinguish tests on dry and saturated samples, respectively)

Peak and residual strength

The repeat tests reveal good reproducibility for the characteristics of the stress–strain curves recorded at the two institutions, further documenting the homogeneity of the blocks

(Table 5). Yet, the standard deviation of repeat tests exceeds the experimental uncertainty for stress difference, suggesting some influence of sample-to-sample variability regarding the distribution of micro-flaws not resolved by bulk properties, such as density or ultrasonic velocity (Table 2).

Table 5 Relative standard deviations (%) of peak strength, residual strength, and Young’s modulus obtained from repeated tests (conditions see Table 3). The number of tested samples is indicated in parenthesis

	Institution	Peak	Residual	Young’s modulus
Crab Orchard sandstone dry	USGS	3.5 (3)	4.3 (3)	4.5 (2)
	RUB	1.3 (3)	10.1 (3)	0.8 (2)
Carrara marble dry	USGS	0.9 (3)*	–	1.0 (2)
	RUB	0.7 (3)*	–	2.1 (3)
Sierra White granite dry	USGS	1.2 (3)	24.7 (3)	0.7 (2)
	RUB	1.6 (5)	15.4 (5)	2.0 (4)
Berea sandstone saturated	USGS	1.1 (3)	1.3 (3)	– (1)
	RUB	0.7 (3)	0.2 (3)	0.7 (3)
Wilkeson sandstone saturated	USGS	0.5 (3)	3.3 (3)	0.8 (2)
	RUB	2.3 (2)	4.7 (2)	8.5 (2)

*Determined at ~2% axial strain, because the repeated tests were performed at conditions, for which Carrara marble exhibits continuous hardening

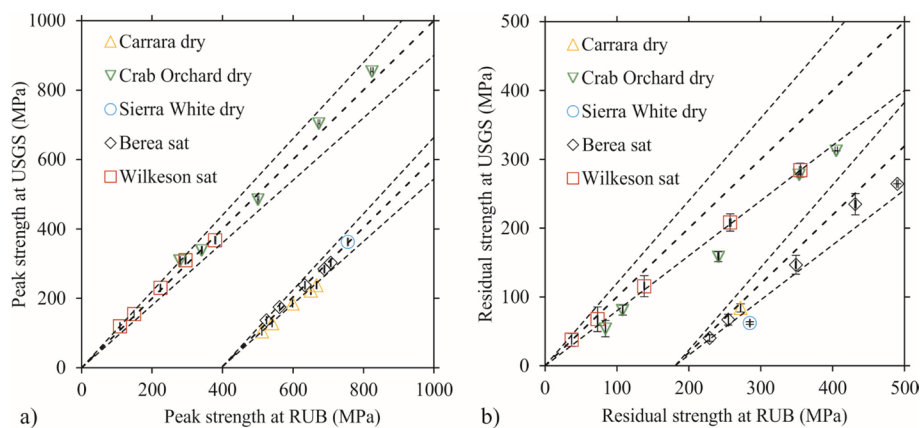


Fig. 3 Comparison between **a** peak strength and **b** residual strength measured at RUB and USGS. In **a**, error bars for RUB indicate the total uncertainty of $\pm 0.4\%$. In **b**, error bars for USGS-data exemplify the strain dependence of residual strength. The long-dashed lines indicate one-to-one identity; the short-dashed lines indicate 10% and 20% deviations for peak and residual strengths, respectively. In

the two plots, the data are split into two groups with results for one shifted along the *x*-axis for presentational purposes. (acronyms USGS and RUB denote data gained at U.S. Geological Survey, Menlo Park, and Ruhr-Universität Bochum, respectively; labels “dry” and “sat” distinguish tests on dry and saturated samples, respectively)

Table 6 Correlation of results for peak strength (USGS vs. RUB, see Fig. 3a) and its uncertainty estimated accounting for the experimental uncertainty of RUB data. (acronyms USGS and RUB denote data

gained at U.S. Geological Survey, Menlo Park, and Ruhr-Universität Bochum, respectively)

Rock type	Slope*	Slope uncertainty*	Slope	Slope uncertainty	Intercept (MPa)	Intercept uncertainty (MPa)
Berea sandstone	1.014	0.005	1.108	0.006	–24.5	1.1
Carrara marble	0.897	0.013	1.203	0.030	–9.8	3.4
Crab Orchard sandstone	1.030	0.005	1.000	0.005	–14.2	2.1
Wilkeson sandstone	1.038	0.004	1.035	0.004	–14.1	0.8

*Regression line forced through the origin

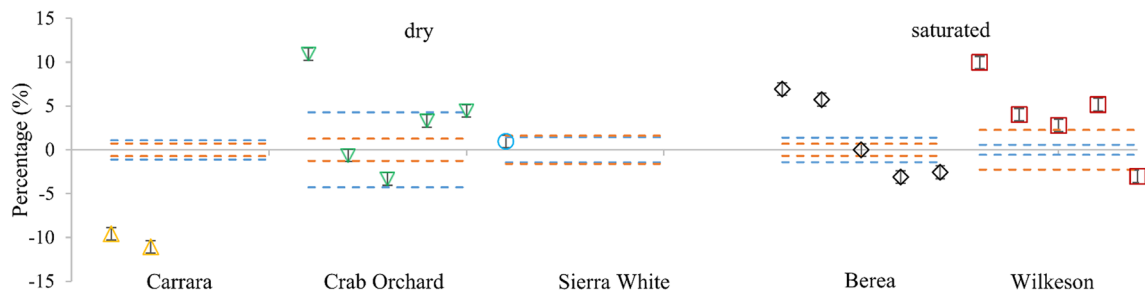


Fig. 4 Deviation of peak-strength values from the identity line (Fig. 3a) in comparison to sample-to-sample variability as derived from standard deviations of the results of repeat tests (dashed lines: blue USGS, orange RUB). A positive deviation indicates that the strength measured by USGS exceeds that measured by RUB. Sym-

bols are plotted in order of increasing confining pressure (see Table 3) from left to right; error bars indicate experimental uncertainty of $\pm 0.4\%$. (acronyms USGS and RUB denote data gained at U.S. Geological Survey, Menlo Park, and Ruhr-Universität Bochum, respectively)

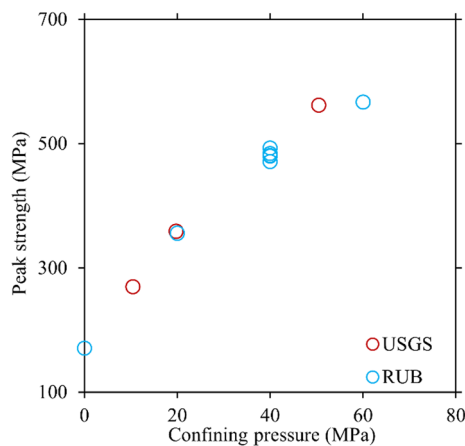


Fig. 5 Peak strength as a function of confining pressure for dry Sierra White granite. (acronyms USGS and RUB denote data gained at U.S. Geological Survey, Menlo Park, and Ruhr-Universität Bochum, respectively)

Peak strengths reported by the two institutions for the suite of rocks span an order of magnitude, with a “weaker” group comprising Carrara marble, Berea and Wilkeson sandstone, and a “stronger” group comprising Crab Orchard sandstone and Sierra White granite, and are generally in close agreement within $< 10\%$ (Fig. 3a), but some systematics in the small deviations are evidenced by the correlation details (Table 6). For all rock types except for Carrara marble, samples tested at USGS appear slightly stronger ($< 12\%$) than those tested at RUB (see also Fig. 4). This observation also applies to Sierra White granite, for which results cannot be fully represented in the cross plots because of differences in the confining pressures applied at the two institutions, judging from a comparison of the trends of strength with pressure (Fig. 5). Unconstrained linear regression between the data sets of the two laboratories leads to intercepts of

a magnitude (Table 6) that we find difficult to plausibly explain by systematic shifts in load measurements or stress determination but attribute to sample-to-sample variability.

The residual strengths determined by USGS tend to be less than the ones determined at RUB for nominally equivalent tests, most notably for Crab Orchard sandstone (Fig. 3b) but also for Sierra White granite (Fig. 5), the two strongest rocks. The effect becomes more significant at higher confining pressures, and is probably partly related to the difference in the extent of overshoot during unstable brittle fracture controlled by the difference in system compliance (Fig. 2a).

Hydraulic permeability

Measured permeability values span approximately six orders of magnitude (Fig. 6). The observed order of magnitude agreement in permeability between the participating laboratories is good considering that four different methods were used. The examination of samples of Berea sandstone and Crab Orchard sandstone drilled in three orthogonal directions by the group at PSU revealed hydraulic anisotropy with the measurement directions of USGS and RUB constituting the least permeable one and the two other directions being up to a factor of two more permeable.

For a single rock, permeability varied up to two orders of magnitude over the explored range in confining pressure. The pressure dependence of permeability differs significantly in two cases. The pressure dependence of Crab Orchard sandstone observed by USGS exceeds that reflected by data from PSU and RUB (Fig. 6b). Berea sandstone did not exhibit a pressure dependence in permeability for the investigated range when tested by the oscillatory method at RUB, while it did for pulse and constant-flux tests performed at PSU (Fig. 6a), albeit with considerable variation during the three loading–unloading cycles (see Appendix C).

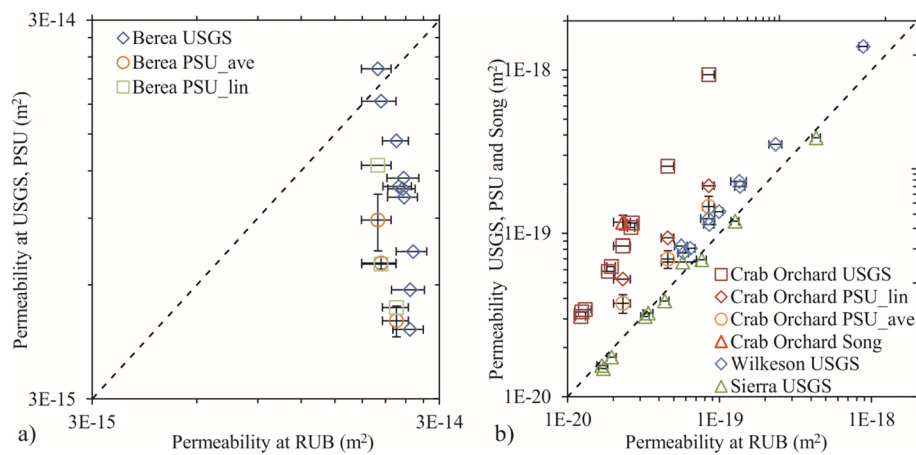


Fig. 6 Comparison between permeability measured at RUB, USGS, and PSU, and by Song et al. (2013), label “Song”. Two calculating methods were used at PSU: average of single tests (avg) and linear approximation (lin). Error bars indicate experimental uncertainty. The dashed line indicates identity. Only results from the first pressurization are plotted (but see Appendix C for the documentation of cycle-

dependence), since the cycling procedure conducted at PSU after the initial pressurization did not fully comply with the recommended test sequence (see Appendix C). (acronyms USGS, PSU, and RUB denote data gained at U.S. Geological Survey, Menlo Park, Penn State University, and Ruhr-Universität Bochum, respectively)

Discussion

As a whole, the results for strength measures confirm that (a) the chosen rocks were suitable for a comparative study, and (b) the accuracies reached by the experimental setups and procedures do not limit the significance of the determined strength measures, in agreement with the conclusions of Pincus (1996). The situation is quite different for the results of the permeability determinations. The consistency between the order of magnitude of results may be considered satisfactory but discrepancies in detail of the results, in particular regarding the pressure dependence of permeability, suggest methodological issues.

Factors affecting deformation characteristics

The slope of a stress–strain curve resulting from a conventional triaxial compression test with a single loading cycle may deviate from the intrinsic static Young’s modulus of the tested material for a number of reasons (Fjær 2019), among them a notable physical one, the irreversible closure of microfractures (David et al. 2020). The accuracy of the transformation of external displacement measurements into sample strain is not only affected by the uncertainty of stiffness calibrations but also by potential tilting owing to non-parallelism of sample and/or piston end faces. The presented apparent moduli provide a way to evaluate the accuracy in strain, relevant, for example, in the light of the determination of characteristic strain values employed as rock-failure criteria (e.g., Aydan et al. 1993; Fujii et al. 1998) and also

for discussions of the mismatch between static and dynamic elastic parameters (e.g., Fjær 2009, 2019).

The values for the apparent static Young’s moduli from the two laboratories fall within the limits expected from the composition of the tested rocks, but only half of them match within 15% with the values determined from tests performed at RUB tending to exceed the ones from tests at USGS. The good correspondence of maximum stress difference (Fig. 3) between the two laboratories suggests that neither uncertainty in stress determination nor imperfect sample geometry can account for the observed trend between the two moduli data sets. The compliances of the assemblies used at USGS and RUB are about 0.002 mm/MPa and 0.001 mm/MPa, respectively, and thus the corrections involved in strain determination amount to up to 70% of the total recorded displacement for tests at USGS on the stiffest rock type, Carrara marble. The compliance calibrations in the two laboratories follow the accepted procedure of testing a steel dummy with supposedly known elastic properties. The discrepancy between the two data sets for static Young’s moduli could well be the result of the successive approximations underlying its determination, i.e., (i) the approximation of the machine compliance by an analytical function used in the correction calculation (USGS: linear, RUB: non-linear) that prominently affects the details of the resulting stress–strain curves in particular during the initial steep increase, and (ii) the degree of the polynomial fit to the pre-peak section of the stress–strain curves. Apart from an overlooked methodological issue, which likely can only be resolved by a round

robin, size dependence may play a role. Observations on size dependence of elastic moduli are not only disparate but also restricted to tests at ambient pressure (e.g., Zhai et al. 2020; Li et al. 2021) and thus may not apply to our set of data from tests at elevated pressure, at which the large microcracks that presumably dominate behavior at ambient pressure are closed.

The compressive strength of brittle materials critically depends on their inventory of microdefects, such as pores and cracks. The suite of tested sandstones serves as an illustrative example for the inverse correlation of strength and porosity. The role of microdefects introduces a random component to strength owing to the variability in the actual realizations of micro-flaw distributions beyond directionally independent bulk properties, such as density. Thus, it is not surprising that strength exhibits a variability beyond measurement accuracy. On average, however, the differences in strength observed for the two institutions are qualitatively and quantitatively in accord with the size-effect of higher strength for smaller samples, commonly considered a consequence of microdefect statistics (e.g., Bernaix 1969; Lockner 1995; Paterson and Wong 2005). For example, a typical strength loss of $\Delta\sigma/\sigma \sim (\Delta L/L)^{-1/2}$ (Lockner 1995) predicts the larger RUB samples to be approximately 8% weaker than the smaller USGS samples. Our results imply that sample size may affect interlaboratory strength comparisons or use of strength data as input in numerical codes. However, we cannot exclude that the differences in preparation contribute to the systematic difference in measured strength. For example, the absence of cylindrical grinding at RUB may facilitate fault nucleation at surface flaws and absolute differences in end-face parallelism between RUB and USGS may cause slight deviations to the stress distribution.

The tests on saturated samples of Wilkeson sandstone were likely not fully drained according to volumetric strain measurements and the constraints on hydraulic diffusivity (Ahrens et al. 2017). Insufficient internal drainage may increase or decrease (or consecutively both depending on the evolution of hydraulic properties during deformation) the effective stress state during deformation, therefore, affecting strength. The shorter samples used by USGS in principle favor effective internal drainage over the longer ones used by RUB. The absence of a substantial difference between the strengths observed in the two laboratories for tests on saturated samples may indicate that the modest length difference does not critically affect internal drainage conditions in this case and/or be related to the generally low dilatancy-hardening potential (Brace and Martin 1968; Duda and Renner 2013) of the experiments performed at a fluid pressure of only 2 MPa. The latter would also

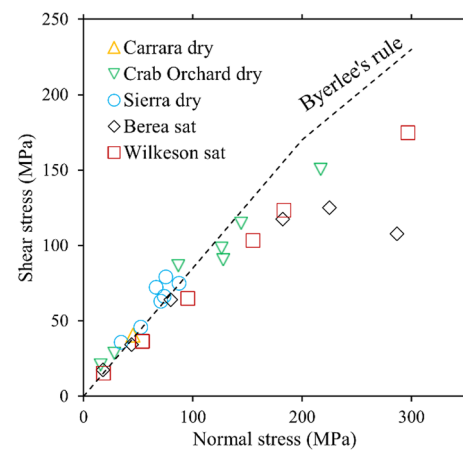


Fig. 7 Normal and shear stresses derived from residual strengths and failure angles observed by Ruhr-Universität Bochum (labels “dry” and “sat” distinguish tests on dry and saturated samples, respectively). The dashed line indicates Byerlee’s bilinear rule (Byerlee 1978)

annihilate possible contributions of differences in design of the interface between sample and piston, i.e., realization of technical drainage, and loading details, e.g., waiting time to reach equilibration after hydrostatic pressurization, and deviatoric loading with constant piston velocity vs. constant strain rate.

Residual strength, in contrast to peak strength, is hard to uniquely determine, because the post-failure section of stress–strain curves typically does not reach a well-defined stress-plateau (Fig. 2a). Ideally, residual strength in brittle faulting represents a constant frictional stress, independent of continued sliding, attained after a fault is fully developed. In practice, sample failure may produce fractures that intersect the loading pistons in contact with the samples or produce fractures with varying fault angles. As a result, reproducibility of residual strength is expected to be worse than for peak strength. Furthermore, the actual contact area of the fracture plane decreases with continued sliding, leading to a decrease in residual stress with increasing axial strain (Fig. 2a), even for a constant friction coefficient. Thus, the difference in absolute strain, at which residual stress was determined, partly controlled by machine stiffness owing to its control on the uncontrolled release of elastic energy stored in the loading pistons in a rapidly failing sample, may account for the difference in residual stress values between the two laboratories. The role of machine stiffness for post-failure characteristics has been noted before (e.g., Hudson et al. 1972; Mansurov 1994); also the jacketing procedure and material as well as sample size may have some effect. Combined with measurements of the shear fracture orientation determined on samples retrieved from the vessel

after the conventional triaxial testing, the residual strengths determined at RUB are in general agreement with Byerlee's rule (Fig. 7) up to about 150 MPa normal stress. Wilkeson sandstone exhibits the lowest friction coefficient, as previously observed for other porous sandstones (Costamagna et al. 2007), in this case possibly related to its fairly large content in phyllosilicates (Table 1; Tembe et al. 2010). The deviations from Byerlee's rule observed for the sandstone samples at normal stresses above about 150 MPa may indicate the increasing contribution of cataclastic flow by pore collapse to their deformation.

Issues related to the determination of hydraulic permeability

Constant flow experiments correspond to the direct implementation of Darcy's law and their results thus exhibit benchmark character for permeability of a specific sample. The analysis procedures of all transient methods assume that samples represent homogeneous and isotropic continua on length scales much smaller than the sample scale, an assumption whose general applicability appears rather debatable in the light of the complexity of the conduit networks of rocks. Nevertheless, Schepp and Renner (2021) showed that constant-flow experiments and oscillatory pore-pressure tests (harmonic pressure-interference) agree within experimental uncertainty for Wilkeson sandstone and Westerly granite, the latter probably a good match for Sierra White granite, when performed on the same sample.

Testing different samples in different laboratories, fundamentally, cannot resolve whether the origin of the differences in permeability results obtained using different methods reflect sample-to-sample variability or methodological characteristics, a limitation that also applies to the recent comparative study of the permeability of Grimsel granodiorite (David et al. 2018a,b). The sample of Crab Orchard sandstone tested by Song et al. (2013) originated from the block used by PSU in this study and has a reported connected porosity of $3.5 \pm 0.1\%$, i.e., almost 2% lower than those tested at RUB (Table 2), pointing to differences between samples from different blocks due to natural variability of the rocks. Yet, the deduced relation in porosity is opposite to the relation in permeability values gained at PSU and RUB (Fig. 6b). Heterogeneity has been demonstrated to be a crucial factor for the outcome of permeability measurements with transient methods, in cases causing a considerable effect of sample size (Song and Renner 2006) that may contribute to the observed differences here, too, owing to the differences in sample diameter used by RUB, and PSU and USGS.

Besides inhomogeneity, anisotropy constitutes an important and yet unresolved issue for permeability determination with transient methods. Judging from the first measurements at the lowest effective pressures performed at PSU in three perpendicular directions, the difference between the most and least permeable direction is less than a factor of 3 for Berea sandstone. The constant-flow tests on samples of Berea sandstone constitute benchmarks for the degree of anisotropy in permeability, possibly including some sample-to-sample variability though. The significance of the anisotropy constraints from constant-head tests on samples of Crab Orchard sandstone, i.e., a ratio of about 2 between least and most permeable direction, however, remains compromised by the unresolved effect of anisotropy on the evaluation strategy. Analytical and/or numerical modeling may facilitate progress in resolving this fundamental problem of the determination of hydraulic properties.

The most significant and suspicious differences in the results for permeability from the three institutions arise from their pressure dependence (Fig. 6), unlikely a result of either heterogeneity or anisotropy of tested samples. The partial convolution of the differences in pressure dependence with significant cycle dependences (Appendix B) may indicate protocol biases involving the actual achievement of pore-pressure equilibration between the various pressure steps, the oscillatory method nominally less depending on equilibration. The systematic inverse correlation of compressive strength of the tested rocks with the differences in pressure dependence and the occurrence of cycle dependence may, however, also indicate a contribution of local failure at sample end-faces in contact with the permeable end-plugs. Dedicated microstructural investigations and design variations could in principle clarify this issue. Finally, differences in the total duration of permeability tests may play a role when the samples contain clay minerals with the potential for swelling, as might be true for Berea sandstone (Table 1).

Conclusions

The sample-to-sample variability inherent to a natural material and the potential size dependence affect the quantitative significance of experimental data from laboratory tests on rock samples for validation of numerical codes. Constraining the actual sample-to-sample variability by basic physical characterization of samples and repeat tests may improve the understanding of the significance of results. Our interlaboratory comparison suggests that

unresolved methodological uncertainties remain for permeability tests and to a much lesser degree for triaxial compression tests that outmatch the error propagation calculations based on the typical accuracy of high-quality sensors used in laboratories by large.

Static Young’s moduli were not included in the “official” work program of the interlaboratory comparison, but we reported results, because the documentation of differences appears instructive regarding the significance of numerical values for this parameter and highlights the importance of clarifying calculation procedures as well as paying attention to machine details, such as the number of external displacement transducers used and the stiffness correction employed. Post-failure more so than failure behavior appears to be an issue of conventional triaxial testing to address further regarding its relation to system stiffness. The interpretation of testing at elevated pore pressure may benefit from a thorough validation of effective drainage conditions.

The results for the various commonly applied methods to determine hydraulic permeability may be affected differently by heterogeneity at the sample scale, and by anisotropy. However, the observed differences in the dependence of permeability on pressure and pressurization history point to the potential benefits of confirming the suitability of the design of apparatus components and of the test procedures. Validation of permeability determinations in the context of digital rock physics (e.g., Mehmani et al. 2020) may have to account for the different boundary conditions used in experiments.

The extensive data set is provided in repositories (Cheng et al. 2023; Lockner et al. 2023) to serve future “benchmarking” intentions, be it to check the performance of new laboratory equipment or of numerical modeling approaches. In particular, the complete records of the deformation tests performed at elevated fluid pressure may allow testing hydro-mechanical codes. A great opportunity to reach progress in the understanding of the role of heterogeneity and anisotropy for laboratory-based constraints on physical properties of rocks lies in the bi-directional exploitation of the synergies between modeling and experimental approaches.

Appendix A

A Overview of comparative studies on mechanical and hydraulic properties of rocks related to the current study (Table 7).

Table 7 Details of comparative studies performed with the aim to assess the significance of measurements of a range of rock-physical properties

Study	# of laboratories	# of materials	# of methods/devices	Determined property	Comment
McPhee and Arthur (1994)	4	1 (Clashach sandstone)		Relative permeability	Round robin
Pincus (1993, 1994)	6 to 8	5 (Barre granite, Berea sandstone, Salem limestone, Tennessee marble, aluminum alloy)		Longitudinal and transverse pulse velocity, uniaxial elastic modulus, indirect tensile strength, unconfined compressive strength	Precision statements presented in accord with ASTM Standard E691 (ASTM, 1999) with repeatability and reproducibility estimated for results from individual laboratories and for all results
Pincus (1996)	6	4 (Barre granite, Berea sandstone, Tennessee marble, aluminum alloy)		Young’s modulus, Poisson’s ratio, ultimate strength	Extends previous study to measurements at elevated confining pressure of 10, 24, 40 MPa
Rasilainen et al., (1996)	4	1 (granitic drill core)	6	Porosity	Highlights problems with low-porosity rocks, i.e., < 1%
Sandström (2006)	2	2 (drillcores from two sites)		Thermal, uniaxial compressive strength, Young’s modulus, tensile strength, residual shear stress	Triaxial compression at 2, 7, 10 MPa confining pressure

Table 7 (continued)

Study	# of laboratories	# of materials	# of methods/devices	Determined property	Comment
David et al., (2018a)	24	1 (Grimsel granodiorite)	3 (transient, steady-state, oscillatory)	Permeability	Difference between test results for gas and liquid; lacks uncertainty analyses
Davy et al. (2019)	4	1 (Coburg limestone)	3 (decay, quasi static, steady-state)	Gas permeability	Emphasize role of sample-to-sample variability due to heterogeneity, and of anisotropy
Ógúnsàmi et al., (2021)	3	1 (sandstone)	3	Elastic properties	True round robin with a single sample
Schepp and Renner (2021)	1	2 (Westerly granite, Wilkeson sandstone)	4 (three oscillatory, steady-state)	Hydraulic	Sample-variability excluded for three methods; emphasize that different pressure-perturbation protocols lead to different investigated sample volumes
focused on low permeability shales					
Ghanizadeh et al. (2015)		1 (Duvernay shale)	3 (profile probing, pulse decay, crushed-rock technique)	Shale permeability	Discrepancies arising from variations in sample size and treatment, stress-state and physical principles of measurement
Profice et al. (2016)	4	2 (shale, pyrophyllite)	3 (step decay, pulse decay, steady state)	Permeability	Round robin for pyrophyllite; role of Klinkenberg correction
Yu et al. (2017)		1 (Opalinus clay)	3 (petrophysical analysis of drill cores, in situ hydraulic tests, tidal analysis)	Hydraulic	In-situ borehole measurements combined with core analyses
Minardi et al. (2021)	3	1 (Opalinus clay)	2 for sample saturation	Undrained triaxial testing	Sample pre-selected based on computer tomography; emphasize roll of consistent sample handling and specimen preparation, and saturation
Winhausen et al. (2021)	1	1 (Opalinus clay)	3 (pore pressure oscillation, pressure pulse decay, pore pressure equilibration)	Hydraulic	Measurements on a single sample over the course of 6 months
Zhang et al. (2022)	1	1 (Eidsvold siltstone)	3 (pressure buildup, pulse decay, steady state)	Permeability	Sample-to-sample variability excluded; possible effect of perturbation history

Appendix B

Details of uncertainty analysis .

A quantitative comparison of experiments from different laboratories performed at the same nominal effective pressure involves determining the uncertainty related to the accuracies of the two pressure gauges measuring the confining and the fluid pressure (we indicate uncertainty of a measured quantity by a leading “ δ ”):

$$\delta p' = \sqrt{\left(\frac{\partial p'}{\partial p_c} \delta p_c\right)^2 + \left(\frac{\partial p'}{\partial p_f} \delta p_f\right)^2} = \sqrt{(\delta p_c)^2 + (\delta p_f)^2}. \tag{1}$$

For the range of pressures typically employed in rock mechanics, sensors are in the accuracy class 0.2, i.e., their readings have an error of $\pm 0.2\%$ of their maximum value, comprising non-linearity, repeatability, and temperature effects, and thus $\delta p_i = 0.2\% \times p_{i,max}$ for $i = c, f$ or

$$\delta p' = 0.2\% \times \sqrt{(p_{c,max})^2 + (p_{f,max})^2}, \tag{2}$$

$$< 0.2\% \times \sqrt{2} \times 150 \text{ MPa} \approx 0.4 \text{ MPa}$$

assuming, for the example calculation, that the range of the two pressure transducers was chosen to match the highest confining pressure employed in this study. The estimate of uncertainty (2) is likely overly conservative, since about half of the error associated with an accuracy class comes from the temperature effect assuming that the operation could be at temperatures deviating as much as $\pm 10 \text{ K}$ from the calibration conditions, while most laboratories will probably have a much lower temperature variation during an experiment.

A stress difference is calculated as

$$\Delta\sigma = \sigma - \sigma_{ref} = \frac{\Delta F}{A}, \tag{3}$$

with a relative uncertainty

$$\frac{\delta\Delta\sigma}{\Delta\sigma} = \sqrt{\left(\frac{\delta\Delta F}{\Delta F}\right)^2 + \left(\frac{\delta A}{A}\right)^2}, \tag{4}$$

where $\sigma, F = F_{ref} + \Delta F$ and A denote the current axial stress load, and cross section of the sample, and σ_{ref} and F_{ref} the reference axial stress and load before deviatoric loading sets in, respectively. For the range of forces typically employed in rock mechanical tests, load cells tend to fall at least in the accuracy classes 0.3 to 0.5 with a relative uncertainty in linearity $\delta F/F|_{lin}$ typically 0.1% or less. Different from effective pressure, representing the difference between two pressures measured with two sensors, the uncertainty of a force difference, determined from two readings of the same instrument within a single loading cycle, results from the non-linearity

of the load cell alone, i.e., $\delta\Delta F/\Delta F = \delta F/F|_{lin}$, unless the difference under consideration is as small as the digital resolution, typically 13 bit or better, depending on the used acquisition system and general noise level. In a triaxial compression test under elevated confining pressure, variable friction might contribute to the uncertainty in force difference. The friction might, for example, increase with the increasing deformation of the axial piston during deviatoric loading. This contribution is difficult to constrain precisely, but an indication of its relevance can be gained from piston cycles at different confining pressures and deviatoric loads. The relative uncertainty in sample cross-sectional stems from the uncertainty of the radius of the prepared sample $\delta A_0/A_0 < 0.3\%$ and the counteracting changes associated with pressurization and deviatoric loading. Stress difference is underestimated at the start of a triaxial test and its increase with axial strain is overestimated when the changes in dimensions of a sample under pressure and axial stress are not accounted for but the initial dimensions are used for the calculation. When ignoring a contribution from variable friction, the relative uncertainty in stress difference is

$$\frac{\delta\Delta\sigma}{\Delta\sigma} = \sqrt{\left(\frac{\delta F}{F}|_{lin}\right)^2 + \left(\frac{\delta A_0}{A_0}\right)^2 + \left(2\nu\varepsilon_{ax} - \frac{2p_c}{3K}\right)^2}. \tag{5}$$

Obviously, the strain dependence of the uncertainty depends on the “elastic” parameters of the sample. The uncertainty in peak stress, typically associated with axial strains of $> 1\%$, is dominated by the changing dimensions of the sample for a bulk modulus $> 10 \text{ GPa}$ and Poisson’s ratios between 0.1 and 0.4.

Axial strain

$$\varepsilon_{ax} = \frac{\Delta d_{corr}}{L}, \tag{6}$$

is deduced from the current displacement of the axial piston, d , corrected for system compliance k_{sys} , e.g., for a linear approach $\Delta d_{corr} = d - k_{sys}\Delta F - d_{HP}$, where d_{HP} denotes the displacement at the hit-point, and the current length of the sample L . The relative uncertainty of axial strain is estimated as

$$\frac{\delta\varepsilon_{ax}}{\varepsilon_{ax}} = \sqrt{\left(\frac{\delta\Delta d_{corr}}{\Delta d_{corr}}\right)^2 + \left(\frac{\delta L}{L}\right)^2}, \tag{7}$$

with

$$\delta\Delta d_{corr} = \sqrt{\left(\frac{\delta d}{d}|_{lin} \Delta d_{corr}\right)^2 + (\delta k_{sys}\Delta F)^2 + (k_{sys}\delta\Delta F)^2}$$

$$= \Delta d_{corr} \sqrt{\left(\frac{\delta d}{d}|_{lin}\right)^2 + \left(\frac{k_{sys}\Delta F}{\Delta d_{corr}}\right)^2 \left[\left(\frac{\delta k_{sys}}{k_{sys}}\right)^2 + \left(\frac{\delta F}{F}|_{lin}\right)^2\right]}. \tag{8}$$

For a typical stiff assembly, the correction makes only a fraction of the corrected value, i.e., $k_{\text{sys}}\Delta F/\Delta d_{\text{corr}} < 1$, and calibration tests may lead to $\delta k_{\text{sys}}/k_{\text{sys}} < 5\%$. In addition, methodological uncertainty arises from the external measurement of displacement related to piston tilting, that may, however, be minimized using three displacement transducers arranged on a circle with a 120° division and averaging their signals. A typical displacement transducer exhibits a non-linearity $\delta d/d|_{\text{lin}} < 0.2\%$ and the uncertainty in current sample length holds $\delta L/L < 2\delta L_0/L_0 \approx 0.02\%$ for a rock sample with a bulk modulus of 5 GPa or larger. Thus, the relative uncertainty in axial strain is actually dominated by the accuracy of the stiffness calibration and may be estimated as $\delta\varepsilon/\varepsilon_{\text{ax}} \approx 1\%$.

Static Young’s moduli are determined from derivative estimates $E = \Delta(\Delta\sigma)/\Delta\varepsilon_{\text{ax}}$ and thus their accuracy

$$\frac{\delta E}{E} \leq \sqrt{\left(\frac{\delta\Delta\sigma}{\Delta(\Delta\sigma)}\right)^2 + \left(\frac{\delta\Delta\varepsilon_{\text{ax}}}{\Delta\varepsilon_{\text{ax}}}\right)^2}, \tag{9}$$

strongly depends on the chosen strain increment $\Delta\varepsilon_{\text{ax}}$. Apart from sensor accuracy considerations, it may be advisable to use increments corresponding to at least 10 times the resolution of the displacement transducer.

Schepp and Renner (2021), and Song et al. (2013) provide extensive uncertainty considerations for constant-rate and oscillatory pore pressure tests, and pulse tests, respectively. For a Darcy test or a constant-rate test, the relative uncertainty in permeability owing to sensor and parameter accuracies amounts to

$$\frac{\delta k}{k}\Big|_{\text{const. rate}} = \sqrt{\left(\frac{\delta\eta}{\eta}\right)^2 + \left(\frac{\delta Q}{Q}\right)^2 + \left(\frac{\delta L}{L}\right)^2 + \left(\frac{\delta A}{A}\right)^2 + \left(\frac{\delta\Delta p}{\Delta p}\right)^2}, \tag{10}$$

where the relative uncertainty in fluid viscosity owing to its temperature and pressure dependence amounts to $\delta\eta/\eta < 10\%$, and the uncertainty in difference between upstream and downstream pressure is calculated analogous to that of effective pressure (1) to

$$\begin{aligned} \delta\Delta p &= \delta(p_{\text{up}} - p_{\text{d}}) \\ &= 0.2\% \times \sqrt{(p_{\text{up,max}})^2 + (p_{\text{d,max}})^2} \\ &< 0.2\% \times \sqrt{2} \times 50 \text{ MPa} \approx 0.1 \text{ MPa}, \end{aligned} \tag{11}$$

where for the numerical example we assumed the use of two identical sensors with a capacity of 50 MPa. When determined from the displacement increments Δd of a pressure intensifier with piston cross section A_p , the uncertainty in flow rate $Q = A_p\Delta d/\Delta t$ results to

$$\frac{\delta Q}{Q} = \sqrt{\left(\frac{\delta A_p}{A_p}\right)^2 + \left(\frac{\delta\Delta d}{\Delta d}\right)^2 + \left(\frac{\delta t}{\Delta t}\right)^2} < 1\%, \tag{12}$$

where the bound estimate holds as long as the displacement increment sufficiently exceeds the resolution of the acquisition system, and for an uncertainty in piston cross section comparable to that quoted above for samples and a non-linearity in displacement transducer of 0.2%. The uncertainty in time δt is in most cases negligible for modern digital acquisition systems, as long as the time interval for the rate determination, Δt , sufficiently exceeds the time step. A linear regression analysis of $\Delta d(t)$ may yield additional uncertainty, e.g., due to temperature fluctuations.

For a pulse-decay test on a sample with a specific storage capacity that is negligible compared to the storage capacities of the upstream reservoir and the downstream reservoir (see Brace et al. 1968), the relative uncertainty in permeability owing to sensor and parameter accuracies amounts to

$$\frac{\delta k}{k}\Big|_{\text{pulse decay}} = \sqrt{\left(\frac{\delta\alpha}{\alpha}\right)^2 + \left(\frac{\delta\eta}{\eta}\right)^2 + \left(\frac{\delta L}{L}\right)^2 + \left(\frac{\delta A}{A}\right)^2 + 2\left(\frac{\delta S}{S}\right)^2}, \tag{13}$$

where $\alpha = \ln \Delta p/\Delta t$ denotes the primary outcome of such a test, the rate of decay of the logarithm of the difference between current and final pressure, and $\delta S/S$ the uncertainty in the involved storage capacities of the two reservoirs, typically $\delta S/S \approx 10\%$. The uncertainty of the decay rate amounts approximately to $\delta\alpha \approx \sqrt{2}\delta p/(\Delta p\Delta t)$, when it is assumed that the uncertainty in time is negligible. Critical issues are accuracy of the pressure difference that is affected by sensor accuracy but also temperature stability in the laboratory, the magnitude of the initially imposed pulse, the finite rise time of the pulse, thermal effects due to the adiabatic heating associated with the pulse, small leaks in the pore pressure system, and the sensitivity of permeability with respect to changes in effective pressure (see Brace et al. 1968). When the specific storage capacity of the sample is of relevant size and to be determined, too, curve fitting of analytical solutions of the pressure diffusion problem is necessitated with involved uncertainty analyses (Song et al. 2013). Uncertainty considerations for constant-head tests are similar to the ones presented here for the pulse-decay method.

For the oscillatory pore-pressure method, the uncertainty in permeability arises from the uncertainty in amplitude ratio and phase shift between downstream and upstream pressure in addition to that in sample geometry and fluid viscosity. The employed sliding-window analysis (Renner and Messar 2006) constrains the uncertainty in the spectral parameters

related to signal stability (e.g., temperature fluctuations) and digital noise. Amplitudes correspond to pressure differences determined with a single sensor, and thus, in addition to the uncertainty gained from spectral analysis, amplitude ratio exhibits an uncertainty determined by the non-linearity of the two pressure sensors:

$$\frac{\delta(\hat{p}_d/\hat{p}_{up})}{(\hat{p}_d/\hat{p}_{up})} = \sqrt{\left(\frac{\delta p_d}{p_d}\bigg|_{lin}\right)^2 + \left(\frac{\delta p_{up}}{p_{up}}\bigg|_{lin}\right)^2} = \sqrt{2} \frac{\delta p}{p}\bigg|_{lin} < 0.2\%, \tag{14}$$

where the second equality holds if two sensors with identical non-linearity are used, assumed to be about 0.1% for the given upper bound that may be severely underestimated if the downstream pressure variation is close to the resolution of the downstream pressure transducer.

Appendix C

Details of pressure and cycle dependence, and variation with preparation direction of permeability estimates (Fig. 8).

Acknowledgements Marc Andre Strutz, Frank Bettenstedt, Nils Güting, Timo Reißner, Timo Stahl, and Benedikt Wöhl are acknowledged for their contributions to the experiments performed at Ruhr-Universität Bochum. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Author contributions Seven authors contributed to this manuscript. David Lockner initiated the interlaboratory study. David Lockner, Joerg Renner, and Demian Saffer organized and oversaw the experiments in the respective laboratories. Mandy Duda, Carolyn Morrow, and Insun Song performed the experiments and processed data. Yan Cheng evaluated data, prepared the figures, and drafted the first version of the manuscript, which was then revised together.

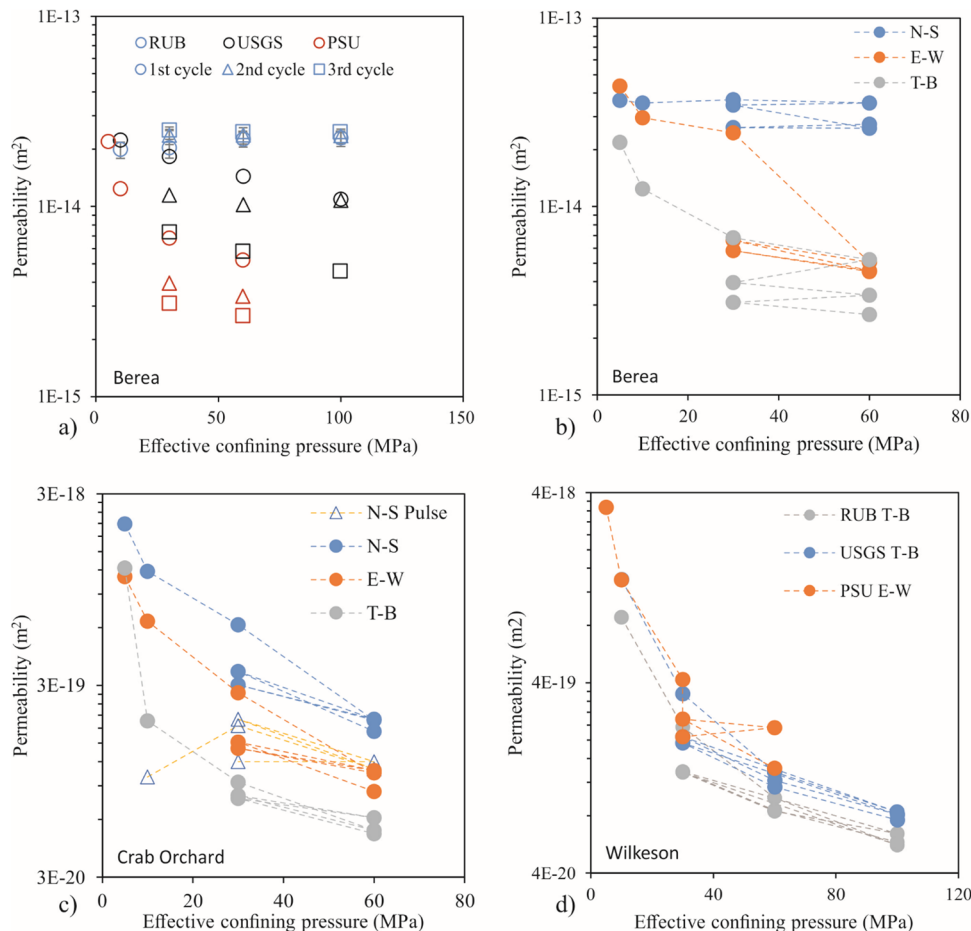
Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The data obtained at the U.S. Geological Survey can be found in Lockner et al. (2023) at <https://doi.org/10.5066/P9WUM58E>. The data obtained at Ruhr-Universität Bochum can be found in Cheng et al. (2023) at <https://doi.org/10.5281/zenodo.8134941>.

Declarations

Competing interests The authors declare no competing interests.

Fig. 8 Pressure and cycle dependence of permeability for a samples of Berea sandstone prepared in uniform direction (T–B, see 2.2 for details of the orientation labelling) and tested at the three laboratories, **b** samples of Berea sandstone prepared in three orthogonal directions at PSU and tested with the constant flow-rate method, **c** samples of Crab Orchard sandstone prepared in three orthogonal directions at PSU and tested with the constant-head method but for the N–S sample that was also tested using the pulse-decay method, and **d** samples of Wilkeson sandstone prepared in T–B direction at USGS and RUB but in E–W direction at PSU. (acronyms USGS, PSU, and RUB denote data gained at U.S. Geological Survey, Menlo Park, Penn State University, and Ruhr-Universität Bochum, respectively)



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahrens B, Duda M, Renner J (2017) Relations between hydraulic properties and ultrasonic velocities during brittle failure of a low-porosity sandstone in laboratory experiments. *Geophys J Int* 212:627–645. <https://doi.org/10.1093/gji/ggx419>
- Almén KE (1994) Exploratory drilling and borehole testing for the nuclear waste disposal programme in Sweden. *Appl Hydrogeol* 2:48–55. <https://doi.org/10.1007/s100400050046>
- ASTM (2017) ASTM D7012-14e1 standard test methods for compressive strength and elastic moduli of intact rock core specimens under varying states of stress and temperatures. American Society for Testing and Materials
- Aydan Ö, Akagi T, Kawamoto T (1993) The squeezing potential of rocks around tunnels; theory and prediction. *Rock Mech Rock Eng* 26:137–163. <https://doi.org/10.1007/bf01023620>
- Benson PM, Meredith PG, Platzman ES, White RE (2005) Pore fabric shape anisotropy in porous sandstones and its relation to elastic wave velocity and permeability anisotropy under hydrostatic pressure. *Int J Rock Mech Min Sci* 42:890–899. <https://doi.org/10.1016/j.ijrmms.2005.05.003>
- Bernabé Y, Mok U, Evans B (2006) A note on the oscillating flow method for measuring rock permeability. *Int J Rock Mech Min Sci* 43:311–316. <https://doi.org/10.1016/j.ijrmms.2005.04.013>
- Bernaix J (1969) New laboratory methods of studying the mechanical properties of rocks. *Int J Rock Mech Min Sci Geomech Abstr* 6:43–90. [https://doi.org/10.1016/0148-9062\(69\)90028-X](https://doi.org/10.1016/0148-9062(69)90028-X)
- Bieniawski ZT, Bernede MJ (1979) Suggested methods for determining the uniaxial compressive strength and deformability of rock materials: Part 1. Suggested method for determining deformability of rock materials in uniaxial compression. *Int J Rock Mech Min Sci Geomech Abstr* 16:138–140. [https://doi.org/10.1016/0148-9062\(79\)91451-7](https://doi.org/10.1016/0148-9062(79)91451-7)
- Birkholzer JT, Bond AE (2022) DECOVALEX-2019: An international collaboration for advancing the understanding and modeling of coupled thermo-hydro-mechanical-chemical (THMC) processes in geological systems. *Int J Rock Mech Min Sci* 154:105097. <https://doi.org/10.1016/j.ijrmms.2022.105097>
- Bossart P (2007) Overview of key experiments on repository characterization in the Mont Terri Rock Laboratory. *Geol Soc Lond Spec Publ* 284:35–40. <https://doi.org/10.1144/SP284.3>
- Brace WF, Martin RJ (1968) A test of the law of effective stress for crystalline rocks of low porosity. *Int J Rock Mech Min Sci Geomech Abstr* 5:415–426. [https://doi.org/10.1016/0148-9062\(68\)90045-4](https://doi.org/10.1016/0148-9062(68)90045-4)
- Brace WF, Walsh JB, Frangos WT (1968) Permeability of granite under high pressure. *J Geophys Res* 73(6):2225–2236
- Byerlee J (1978) Friction of rocks. *Pure Appl Geophys* 116:615–626. <https://doi.org/10.1007/BF00876528>
- Cheng Y, Duda M, Renner J (2023) Interlaboratory comparison of testing hydraulic, elastic, and failure properties in compression: data from Ruhr-Universität Bochum. <https://doi.org/10.5281/zenodo.8134941>
- Churcher PL, French PR, Shaw JC, Schramm LL (1991) Rock properties of Berea sandstone, baker dolomite, and Indiana limestone. In: SPE International Symposium on Oilfield Chemistry. Society of Petroleum Engineers, p 19
- Costamagna R, Renner J, Bruhns OT (2007) Relationship between fracture and friction for brittle rocks. *Mech Mater* 39:291–301. <https://doi.org/10.1016/j.mechmat.2006.06.001>
- David C, Wassermann J, Amann F et al (2018a) KG²B, a collaborative benchmarking exercise for estimating the permeability of the Grimsel granodiorite—Part 1: measurements, pressure dependence and pore-fluid effects. *Geophys J Int* 215:799–824. <https://doi.org/10.1093/gji/ggy304>
- David C, Wassermann J, Amann F et al (2018b) KG²B, a collaborative benchmarking exercise for estimating the permeability of the Grimsel granodiorite—Part 2: modelling, microstructures and complementary data. *Geophys J Int* 215:825–843. <https://doi.org/10.1093/gji/ggy305>
- David EC, Brantut N, Hirth G (2020) Sliding crack model for non-linearity and hysteresis in the triaxial stress-strain curve of rock, and application to antigorite deformation. *J Geophys Res Solid Earth* 125:e2019. <https://doi.org/10.1029/2019JB018970>
- Davy C, Hu Z, Selvadurai P, et al (2019) Transport properties of the Cobourg limestone: a benchmark investigation
- Deb P, Salimzadeh S, Vogler D et al (2021) Verification of coupled hydraulic fracturing simulators using laboratory-scale experiments. *Rock Mech Rock Eng* 54:2881–2902. <https://doi.org/10.1007/s00603-021-02425-y>
- Delay J, Rebours H, Vinsot A, Robin P (2007) Scientific investigation in deep wells for nuclear waste disposal studies at the Meuse/Haute Marne underground research laboratory, Northeastern France. *Phys Chem Earth Parts ABC* 32:42–57. <https://doi.org/10.1016/j.pce.2005.11.004>
- Diehl P, Prudhomme S, Lévesque M (2019) A review of benchmark experiments for the validation of peridynamics models. *J Peridynamics Nonlocal Model* 1:14–35. <https://doi.org/10.1007/s42102-018-0004-x>
- Duda M, Renner J (2013) The weakening effect of water on the brittle failure strength of sandstone. *Geophys J Int* 192:1091–1108. <https://doi.org/10.1093/gji/ggs090>
- Esterhuizen GS (2014) Extending empirical evidence through numerical modelling in rock engineering design. *J South Afr Inst Min Metall* 114:755–764
- Fjaer E (2009) Static and dynamic moduli of a weak sandstone. *Geophysics* 74:103–112. <https://doi.org/10.1190/1.3052113>
- Fjær E (2019) Relations between static and dynamic moduli of sedimentary rocks. *Geophys Prospect* 67:128–139. <https://doi.org/10.1111/1365-2478.12711>
- Fridleifsson GO, Elders WA (2005) The Iceland deep drilling project: a search for deep unconventional geothermal resources. *Geothermics* 34:269–285. <https://doi.org/10.1016/j.geothermics.2004.11.004>
- Fujii Y, Kiyama T, Ishijima Y, Kodama J (1998) Examination of a rock failure criterion based on circumferential tensile strain. *Pure Appl Geophys* 152:551–577. <https://doi.org/10.1007/s000240050167>
- Ghanizadeh A, Bhowmik S, Haeri-Ardakani O et al (2015) A comparison of shale permeability coefficients derived using multiple non-steady-state measurement techniques: examples from the Duvernay formation, Alberta (Canada). *Fuel* 140:371–387. <https://doi.org/10.1016/j.fuel.2014.09.073>

- Ghassemi A (2012) A review of some rock mechanics issues in geothermal reservoir development. *Geotech Geol Eng* 30:647–664. <https://doi.org/10.1007/s10706-012-9508-3>
- Heap MJ, Violay MES (2021) The mechanical behaviour and failure modes of volcanic rocks: a review. *Bull Volcanol* 83:33. <https://doi.org/10.1007/s00445-021-01447-2>
- Holt RM, Kjølås J, Larsen I et al (2005) Comparison between controlled laboratory experiments and discrete particle simulations of the mechanical behaviour of rock. *Int J Rock Mech Min Sci* 42:985–995. <https://doi.org/10.1016/j.ijrmms.2005.05.006>
- Hudson JA, Crouch SL, Fairhurst C (1972) Soft, stiff and servo-controlled testing machines: a review with reference to rock failure. *Eng Geol* 6:155–189. [https://doi.org/10.1016/0013-7952\(72\)90001-4](https://doi.org/10.1016/0013-7952(72)90001-4)
- ISO/IEC (2017) General requirements for the competence of testing and calibration laboratories. In: ISO/IEC17025:2017(en). <https://www.iso.org/obp/ui/#iso:std:iso-iec:17025:ed-3:v1:en>. Accessed 1 Jan 2023
- Jing L (2003) A review of techniques, advances and outstanding issues in numerical modelling for rock mechanics and rock engineering. *Int J Rock Mech Min Sci* 40:283–353. [https://doi.org/10.1016/S1365-1609\(03\)00013-3](https://doi.org/10.1016/S1365-1609(03)00013-3)
- Kolditz O, Fischer T, Frühwirth T et al (2021) GeomInt: geomechanical integrity of host and barrier rocks—experiments, models and analysis of discontinuities. *Environ Earth Sci* 80:509. <https://doi.org/10.1007/s12665-021-09787-0>
- Kovari K, Tisa A, Einstein HH, Franklin JA (1983) Suggested methods for determining the strength of rock materials in triaxial compression: revised version. *Int J Rock Mech Min Sci Geomech Abstr* 20:285–290. [https://doi.org/10.1016/0148-9062\(83\)90598-3](https://doi.org/10.1016/0148-9062(83)90598-3)
- Li Q, Chen L, Sui Z et al (2019) Dynamic analysis and criterion evaluation on rockburst considering the fractured dissipative energy. *Adv Mech Eng* 11:1687814019825906. <https://doi.org/10.1177/1687814019825906>
- Li H, Song K, Tang M et al (2021) Determination of scale effects on mechanical properties of Berea sandstone. *Geofluids* 2021:6637371. <https://doi.org/10.1155/2021/6637371>
- Lockner DA (1995) Rock failure. In: Ahrens TJ (ed) *Rock physics & phase relations*. American Geophysical Union, pp 127–147
- Lockner DA (1998) A generalized law for brittle deformation of Westerly granite. *J Geophys Res Solid Earth* 103:5107–5123. <https://doi.org/10.1029/97jb03211>
- Lockner DA, Marone CJ, Saffer D (2009) SAFOD interlaboratory test, a progress report (abstract). In: *EarthScope 2009 National Meeting*. Boise
- Lockner DA, Cheng Y, Duda M et al (2023) Data for the manuscript: Interlaboratory comparison of testing hydraulic, elastic, and failure properties in compression. U.S. Geological Survey data release. <https://doi.org/10.5066/P9WUM58E>
- Logan JM, Marone CJ, Lockner DA (2010) Inter-lab strength and friction correlations on SAFOD samples (abstract). In: *Supplemental issue Fall AGU Mtg. EOS, Trans. American Geophys. Union*
- Mansurov VA (1994) Acoustic emission from failing rock behaviour. *Rock Mech Rock Eng* 27:173–182. <https://doi.org/10.1007/bf01020309>
- McPhee CA, Arthur KG (1994) Relative permeability measurements: an inter-laboratory comparison. In: *European Petroleum Conference*. SPE-28826-MS
- Mehmani A, Kelly S, Torres-Verdín C (2020) Leveraging digital rock physics workflows in unconventional petrophysics: A review of opportunities, challenges, and benchmarking. *J Pet Sci Eng* 190:107083. <https://doi.org/10.1016/j.petrol.2020.107083>
- Miller SA, Florence AL (1991) Laboratory particle-velocity experiments on Indiana limestone and Sierra white granite. Final report, 5 Oct 90–Jan 92
- Minardi A, Giger SB, Ewy RT et al (2021) Benchmark study of undrained triaxial testing of Opalinus Clay shale: results and implications for robust testing. *Geomech Energy Environ* 25:100210. <https://doi.org/10.1016/j.gete.2020.100210>
- Neuzil CE (2003) Hydromechanical coupling in geologic processes. *Hydrogeol J* 11:41–83. <https://doi.org/10.1007/s10040-002-0230-8>
- Ògúnsàmi A, Jackson I, Borgomano JVM et al (2021) Elastic properties of a reservoir sandstone: a broadband inter-laboratory benchmarking exercise. *Geophys Prospect* 69:404–418. <https://doi.org/10.1111/1365-2478.13048>
- Paterson MS, Wong T (2005) *Experimental rock deformation—the brittle field*. Springer-Verlag, Berlin
- Pincus HJ (1993) Interlaboratory testing program for rock properties (ITPRP): round one. longitudinal and transverse pulse velocities, unconfined compressive strength, uniaxial elastic modulus, and splitting tensile strength. *Geotech Test J* 16:138–163
- Pincus HJ (1994) Addendum to interlaboratory testing program for rock properties, round one. *Geotech Test J* 17:256–258
- Pincus HJ (1996) Interlaboratory testing program for rock properties, round two—confined compression: Young’s modulus, Poisson’s ratio, and ultimate strength. *Geotech Test J* 19:321–336
- Prior DB, Doyle EH (1984) Geological hazard surveying for exploratory drilling in water depths of 2000 meters. In: *Offshore Technology Conference*. OTC-4747-MS
- Profice S, Hamon G, Nicot B (2016) Low-permeability measurements: insights. *Petrophys SPWLA J Form Eval Reserv Descr* 57:30–40
- Rasilainen K, Hellmuth KH, Kivekaes L, et al (1996) An interlaboratory comparison of methods for measuring rock matrix porosity. Finland
- Renner J, Messar M (2006) Periodic pumping tests. *Geophys J Int* 167:479–493. <https://doi.org/10.1111/j.1365-246X.2006.02984.x>
- Sander R, Pan Z, Connell LD (2017) Laboratory measurement of low permeability unconventional gas reservoir rocks: a review of experimental methods. *J Nat Gas Sci Eng* 37:248–279. <https://doi.org/10.1016/j.jngse.2016.11.041>
- Sandström M (2006) Forsmark and Oskarshamn site investigation Borholes Inter-laboratory comparison of rock mechanics testing results. SKB P-05–239
- Schep LL, Renner J (2021) Evidence for the heterogeneity of the pore structure of rocks from comparing the results of various techniques for measuring hydraulic properties. *Transp Porous Media* 136:217–243. <https://doi.org/10.1007/s11242-020-01508-8>
- Schmid SM, Paterson MS, Boland JN (1980) High temperature flow and dynamic recrystallization in Carrara marble. *Tectonophysics* 65:245–280. [https://doi.org/10.1016/0040-1951\(80\)90077-3](https://doi.org/10.1016/0040-1951(80)90077-3)
- Song I, Renner J (2006) Experimental investigation into the scale dependence of fluid transport in heterogeneous rocks. *Pure Appl Geophys* 163:2103–2123. <https://doi.org/10.1007/s00024-006-0121-3>
- Song I, Renner J (2007) Analysis of oscillatory fluid flow through rock samples. *Geophys J Int* 170:195–204. <https://doi.org/10.1111/j.1365-246X.2007.03339.x>
- Song I, Rathbun AP, Saffer DM (2013) Uncertainty analysis for the determination of permeability and specific storage from the pulse-transient technique. *Int J Rock Mech Min Sci* 64:105–111. <https://doi.org/10.1016/j.ijrmms.2013.08.032>
- Tembe S, Lockner DA, Wong T-F (2010) Effect of clay content and mineralogy on frictional sliding behavior of simulated gouges: Binary and ternary mixtures of quartz, illite, and montmorillonite.

- J Geophys Res Solid Earth. <https://doi.org/10.1029/2009JB006383>
- Vanorio T, Di Bonito C, Clark AC (2014) A virtual rock physics laboratory through visualized and interactive experiments. In: AGU Fall Meeting Abstracts. pp ED52A-06
- Wang J, Apel DB, Pu Y et al (2021) Numerical modeling for rockbursts: a state-of-the-art review. *J Rock Mech Geotech Eng* 13:457–478. <https://doi.org/10.1016/j.jrmge.2020.09.011>
- Winhausen L, Amann-Hildenbrand A, Fink R et al (2021) A comparative study on methods for determining the hydraulic properties of a clay shale. *Geophys J Int* 224:1523–1539. <https://doi.org/10.1093/gji/ggaa532>
- Yu C, Matray J-M, Gonçalves J et al (2017) Comparative study of methods to estimate hydraulic parameters in the hydraulically undisturbed Opalinus Clay (Switzerland). *Swiss J Geosci* 110:85–104. <https://doi.org/10.1007/s00015-016-0257-9>
- Zhai H, Masoumi H, Zoorabadi M, Canbulat I (2020) Size-dependent behaviour of weak intact rocks. *Rock Mech Rock Eng* 53:3563–3587. <https://doi.org/10.1007/s00603-020-02117-z>
- Zhang X, Xia Y, Zeng G et al (2018) Numerical and experimental investigation of rock breaking method under free surface by TBM disc cutter. *J Cent South Univ* 25:2107–2118. <https://doi.org/10.1007/s11771-018-3900-y>
- Zhang D, Gao H, Ranjith PG et al (2022) Experimental and theoretical study on comparisons of some gas permeability test methods for tight rocks. *Rock Mech Rock Eng* 55:3153–3169. <https://doi.org/10.1007/s00603-022-02813-y>
- Zhu W, Ougier-simonin A, Lisabeth HP, Banker JS (2012) Developing a virtual rock deformation laboratory. In: AGU Fall Meeting Abstracts. pp ED51A-0870
- Zoback M, Hickman S, Ellsworth W (2010) Scientific drilling into the San Andreas fault zone. *Eos Trans Am Geophys Union* 91:197–199. <https://doi.org/10.1029/2010eo220001>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.