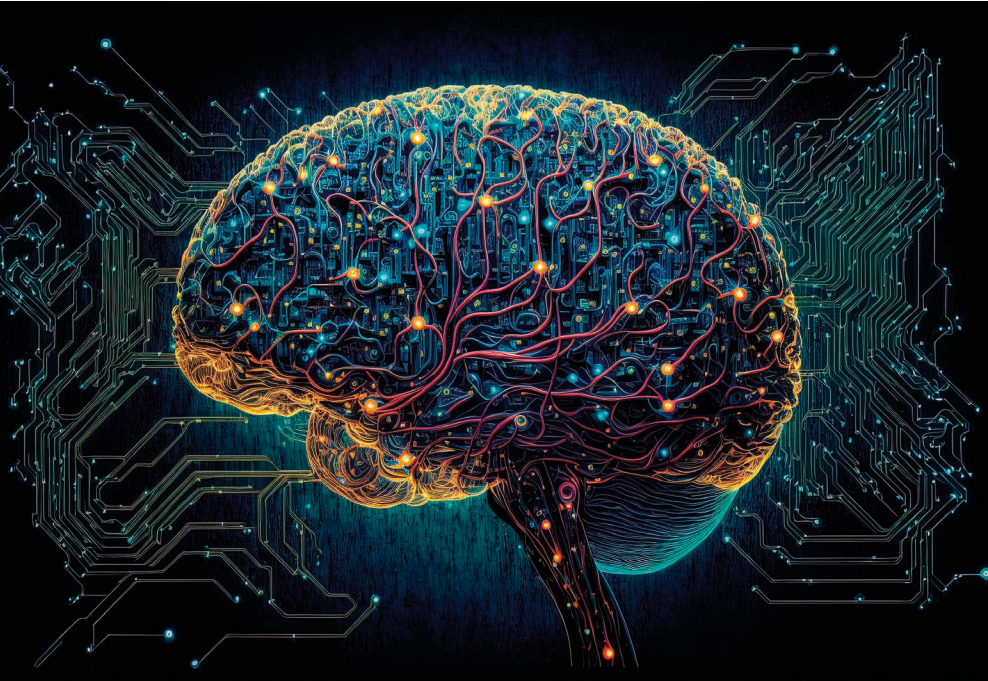


Durchbruch in der neuronalen Telemetrie

Neuer Chip ermöglicht eine zehnfache Datenkomprimierung bei gleichbleibender Signalintegrität



Die neuromorphe Kompressionstechnologie bildet die Grundlage für skalierbare Gehirn-Computer-Schnittstellen (BCIs).

Die Suche nach neuen Wegen zur Behandlung kognitiver, sensorischer und motorischer Störungen sowie der damit verbundenen Beeinträchtigungen – von der Wiederherstellung der Bewegungsfähigkeit bei Menschen mit Lähmungen über die intuitive Steuerung von Prothesen bis hin zur Wiederherstellung von Sprache und Sehvermögen – läuft auf Hochtouren. Parallel dazu drängt die Neurowissenschaft auf leistungsfähigere Werkzeuge, um neuronale Dynamiken zu untersuchen und die Mechanismen zu entschlüsseln, die der Kognition zugrunde liegen. Diese Fortschritte werden zunehmend von Gehirn-Computer-Schnittstellen (BCIs) vorangetrieben, die das Gehirn direkt mit elektronischen Systemen verbinden und sowohl für transformative Therapien als auch für fundiertere wissenschaftliche Erkenntnisse enorme Chancen bieten.

Kortikale BCIs

Systeme, die die elektrische Aktivität der Großhirnrinde erfassen – gibt es in verschiedenen Ausführungen. Intrakortikale BCIs (iBCIs) beispielsweise nutzen Mikroelektrodenarrays (MEAs), die in die Großhirnrinde implantiert werden, während bei auf Elektrokortikographie (ECoG) basierenden Systemen Elektroden auf der Oberfläche der Großhirnrinde, zwischen Schädel und Hirngewebe, angebracht werden. Trotz ihrer Unterschiede zielen beide darauf ab, fein aufgelöste elektrische Aktivität von großen Neuronenpopulationen zu erfassen. Doch mit der steigenden Anzahl der zugrunde liegenden Aufzeichnungskanäle wächst auch das Volumen der neuronalen Daten, die übertragen und verarbeitet werden müssen.

Die Herausforderung

Diese Datenflut treibt den Energieverbrauch und damit die Wärmeentwicklung in die Höhe. Dabei können schon geringe Temperaturanstiege Neuronen irreversibel schädigen. Folglich werden verlustfreie Datenreduktion und -komprimierung unerlässlich, um die Anzahl der übertragenen Bits zu verringern, ohne die Genauigkeit der zugrunde liegenden neuronalen Informationen zu beeinträchtigen.

Skalierung von BCIs

Warum es nicht ausreicht, „einfach nur“ die Anzahl der (MEA-)Aufzeichnungskanäle und die

(drahtlose) Bandbreite zu erhöhen: Die Skalierung von BCIs ist eine äußerst komplexe Herausforderung auf Systemebene. Nehmen wir iBCIs als Beispiel. Zunächst müssen auf der Erfassungsseite die MEA-Neuralsonden hinsichtlich der Kanalanzahl weiter ausgebaut werden, um langfristig mehrere tausend parallele Elektroden zu realisieren – weit über die 1.536 Kanäle hinaus, die das derzeitige (Stand-of-the-Art-)System „Neuropixels 2.0 Quad Base (QB)“ unterstützt.

Am anderen Ende des Systems müssen (i) BCIs eine Kommunikation mit hoher Bandbreite und geringer Latenz zu (externen) Dekodierungs- und Verarbeitungshubs aufrechterhalten. Hier hat sich Impulsfunk-Ultrabreitband (IR-UWB) als vielversprechende Technologie herauskristallisiert. IR-UWB beseitigt nicht nur die Einschränkungen hinsichtlich Benutzerfreundlichkeit und Komfort, die bei kabelgebundenen Verbindungen bestehen, sondern kombiniert auch die Einhaltung elektromagnetischer Vorschriften mit Datenraten von über 124 Mbit/s über Entfernungen von einigen zehn bis mehreren hundert Zentimetern, geringem Stromverbrauch (rund 30 mW, was etwa zehnmal niedriger ist als bei WLAN), hoher Störfestigkeit und einer inhärenten Sicherheit auf der physikalischen Ebene.

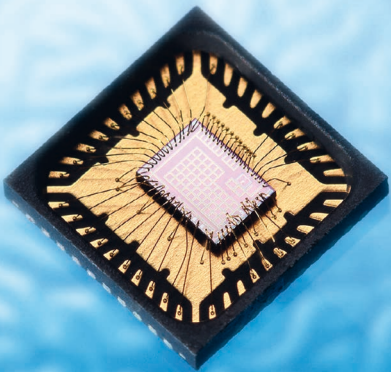
Hohe Bandbreitenanforderungen

Dennoch können selbst die fortschrittlichsten UWB-Verbindungen die Bandbreitenanforderungen künftiger MEAs mit hoher Dichte nicht erfüllen. Das Streamen von Rohdaten von einer bestehenden 1.500-Kanal-Sonde wie dem Neuropixels 2.0 QB erfordert Durchsätze von weit über 500 Mbit/s, was weit über den praktisch nutzbaren Betriebsbereich von UWB hinausgeht. Der Trend zu mehr als 10.000 parallelen Kanälen vergrößert diese Lücke nur noch weiter.

Diese Engpässe verlagern den Druck auf die On-Chip-Kompressionstechnologie, die die MEAs (oder andere Aufzeichnungsgeräte) mit der drahtlosen Schnittstelle verbindet. Konkret muss sie fortschrittliche, verlustfreie Datenreduktion integrieren, um das Datenvolumen drastisch zu verringern und gleichzeitig den vollen Dynamikbereich und den Informationsgehalt der aufgezeichneten Signale zu bewahren. Leider stützen sich herkömmliche Strategien auf große Speicherpuffer, aufwendige digitale Logik oder verlustbehaftete Näherungen – was sie für den Einsatz in stark begrenzten (i)BCIs ungeeignet macht.

Neuromorphe kompressive Telemetrie

Ein Chip für neuromorphe kompressive Telemetrie (NCT) zur verlustfreien Datenreduktion: Um die Anforderungen hinsichtlich Datenrate, Stromverbrauch und Wärmeentwicklung bei (intrakortikalen) BCIs der nächsten Generation



„Send-on-Delta“- Abtast-/Kodierungsansatz

Der „Send-on-Delta“-Abtast-/Kodierungsansatz von Imec verfolgt einen grundlegend anderen Weg. Anstelle einer Abtastung in festen Intervallen schlägt „Send-on-Delta“ ein ereignisbasiertes, signalabhängiges zeitliches Abtastschema vor: Daten werden nur dann generiert, wenn sich ein Signal um mehr als einen vordefinierten Schwellenwert (Δ) ändert. Somit ist die Ausgabe keine dichte Wellenform, sondern ein dünner Strom informationsreicher Ereignisse.

Dies bringt mehrere Vorteile mit sich: drastisch weniger Datenpunkte (oft um einen Faktor von zehn), einen deutlich geringeren Stromverbrauch und einen wesentlich geringeren Bandbreitenbedarf – dabei werden alle Signalspitzen mit hoher Genauigkeit erfasst.

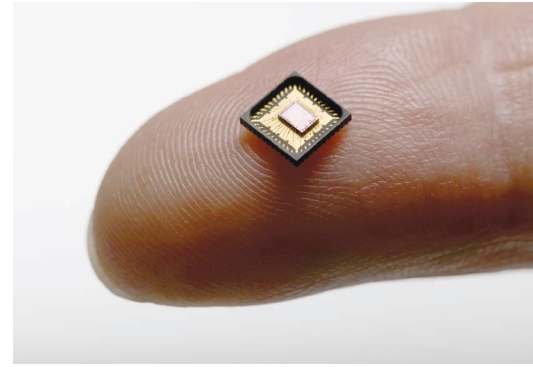
Digitale Codierung

Eine wesentliche Verbesserung im neuesten „Send-on-Delta“-Mechanismus von Imec (zweite Generation) besteht darin, dass die Codierung nun vollständig im digitalen Bereich erfolgt: Anstatt von rohen analogen Spannungen auszugehen, die einen stromhungrigen „Send-on-Delta“-Analog-Digital-Wandler (ADC) durchlaufen, arbeitet das System mit einer Darstellung digitaler Zustände, die aussagekräftige Änderungen im neuronalen Signal widerspiegelt. Einfach ausgedrückt erkennt „Send-on-Delta“ digital, wann sich das Signal ändert, und entscheidet dann, was mit den zugrunde liegenden Daten geschehen soll.

Paketbasiertes AER-Protokoll

Ein ternäres, paketbasiertes AER-Protokoll für erweiterte Paketzusammenstellung und Serialisierung: Der „Send-on-Delta“-Ansatz von Imec nutzt zwar die Sparsamkeit der neuronalen Aktivität effektiv aus, erzeugt jedoch naturgemäß spike-gesteuerte Datenströme (nur bei Änderungen der neuronalen Signale, nicht in festen Intervallen). Dies ist zwar wünschenswert, um Energie einzusparen, erfordert jedoch eine Kommunikationsmethode, die mit unregelmäßigen, spike-gesteuerten Daten umgehen kann.

Address-Event-Representation-Protokolle (AER) sind eine gängige Lösung für die spike-gesteuerte Ereigniskommunikation. Bestehende AER-Verfahren weisen jedoch bei der Anwendung auf neuronale Aufzeichnungen mit hoher Dichte mehrere Einschränkungen auf. Wenn beispielsweise mehrere Auslesekanäle gleichzeitig Ereignisse generieren, stützt sich klassisches AER auf Ereignisarbitrierung oder bestätigungsbasiertes Handshaking – was sich nicht gut auf eine große Anzahl von Kanälen skalieren lässt und unvorhersehbare Latenzen mit sich bringt. Darüber hinaus weisen neuronale Spikes eine starke räumliche Korrelation auf – ein einzelner Spike kann über mehrere benachbarte Elektroden hinweg auftreten –



doch traditionelle AER-Methoden paketieren und serialisieren jedes Ereignis unabhängig voneinander, wodurch wiederholt redundante Adressinformationen übertragen werden und unnötiger Protokoll-Overhead entsteht.

Ereignisbasierter Serializer

Um diese Beschränkungen zu überwinden, hat Imec einen ereignisbasierten Serializer (eSER) entwickelt, der „Send-on-Delta“ mit einem ternären, paketbasierten AER-Protokoll kombiniert – speziell für die neuronale Telemetrie konzipiert. Das Design von Imec bietet mehrere entscheidende Vorteile:

- Ereignisgesteuerte serielle Übertragung – nur wenn neuronale Aktivität stattfindet
- Räumliche Gruppierung korrelierter Ereignisse – das Senden eines einzigen kompakten Pakets anstelle vieler kleiner Nachrichten, wodurch redundante Metadaten vermieden und der Protokoll-Overhead auf bis auf die Hälfte reduziert wird
- Keine Arbitrierungs- oder Kollisionsbehandlungslogik erforderlich – anstatt zwischen gleichzeitigen Ereignissen zu vermitteln, sammelt der eSER zunächst alle Δ -Ausgänge und sendet dann ein Paket in einer geregelten Reihenfolge aus; dadurch werden Ereigniskollisionen vollständig vermieden, während gleichzeitig komplexe -Arbitrierungsschaltungen mit unbestimmter Latenz entfallen – ein wesentlicher Engpass bei herkömmlichen AER
- Umfassende Multi-Bit-Kodierung (ternär) für verlustfreie Rekonstruktion – Die AER-Pakete von Imec enthalten Δ -Werte, die Änderungsrichtung und die Kanal-ID, um eine verlustfreie Rekonstruktion der Spike-Wellenform zu ermöglichen (selbst bei Spikes mit geringer Amplitude bis hinunter zu $\sim 31 \mu\text{V}$)

Somit löst das AER von Imec die Probleme traditioneller Implementierungen hinsichtlich Skalierbarkeit, Komplexität, Overhead, (indeterministischer) Latenz und Stromverbrauch, indem es die Kommunikation an die Eigenschaften neuronaler Signale anpasst – diese sind selten, burstartig und räumlich korreliert. Durch die intelligente Gruppierung von Ereignissen, die Kodierung umfangreicherer Δ -Informationen und die

zu erfüllen, hat Imec einen neuen Chip für neuromorphe Kompressionstelemetrie (NCT) entwickelt, der eine verlustfreie Datenreduktion in Echtzeit ermöglicht. Die Architektur basiert auf zwei zentralen Innovationen (Y. He et al., 2024), (P. Russo et al., 2026):

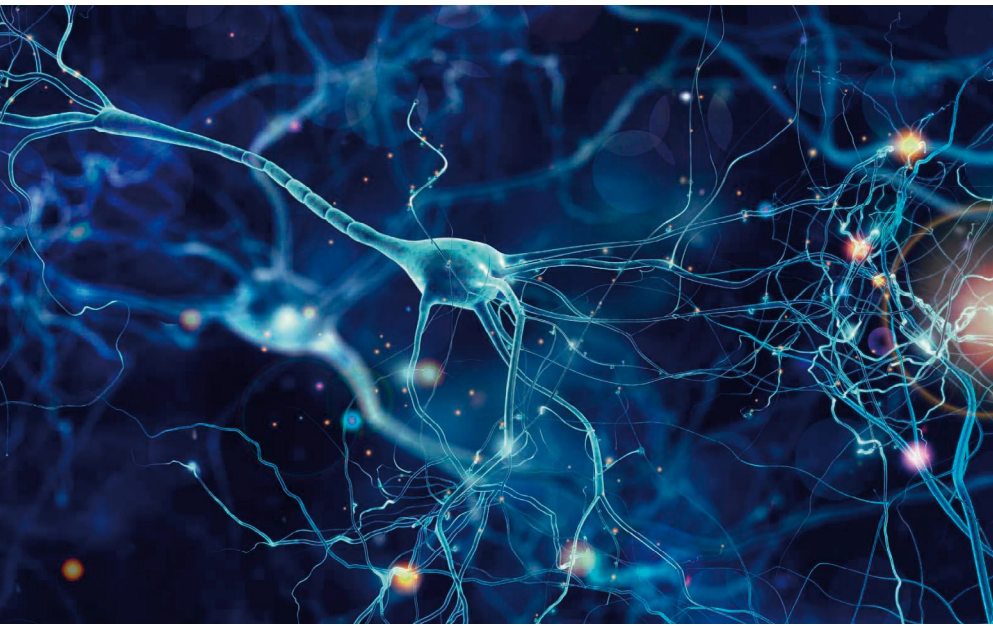
- Erfassung von Send-on-Delta-Signalen, bei der die herkömmliche Abtastung mit Nyquist-Rate durch ein ereignisgesteuertes Verfahren ersetzt wird, das Daten nur dann liefert, wenn sich das neuronale Signal ändert.
- Ein dreifacher paketbasierter AER-Serializer (eSER), der diese Ereignisse zu kompakten Paketen bündelt, um eine effiziente Serialisierung und eine deterministische Übertragung zu gewährleisten.

Zusammen ermöglichen diese Bausteine dem NCT, redundante Daten zu eliminieren – wodurch der Energie- und Bandbreitenbedarf von (i)BCIs gesenkt wird, während gleichzeitig alle für eine originalgetreue Rekonstruktion der Aktionspotenziale erforderlichen Informationen erhalten bleiben.

Send-on-Delta-Kodierung

für die verlustfreie, ereignisgesteuerte Signalerfassung: Die meisten kortikalen Neuronen feuern überraschend selten – typischerweise mit weniger als zehn Hertz, was nur wenigen Dutzend Spikes pro Sekunde entspricht (und oft sogar noch weniger). Diese inhärente Seltenheit bietet eine große Chance für die Datenkomprimierung und -reduktion.

Bei der herkömmlichen Abtastung mit Nyquist-Rate werden Signale mit einer festen Frequenz erfasst – üblicherweise 20 - 30 kHz für die neuronale Erfassung –, unabhängig davon, ob tatsächlich ein neuronales Ereignis (d. h. ein Impuls) auftritt. Dies erzeugt einen kontinuierlichen Strom von Abtastwerten, von denen die überwiegende Mehrheit redundant ist (wenn Neuronen inaktiv sind).



Aktivierung des Serializers nur bei Bedarf (wenn Δ ungleich Null ist) filtert das System redundante Daten bereits an der Quelle heraus und erzielt eine deutlich höhere Komprimierung sowie einen extrem stromsparenden Betrieb.

Validierung

der hochpräzisen, stromsparenden Telemetrie von imec anhand neuronaler Aufzeichnungen: Um die Leistungsfähigkeit zu bewerten, testete imec seinen NCT-Chip – gefertigt in 65-nm-CMOS-Technologie – anhand von in-vivo-Aufzeichnungen neuronaler Aktivität aus Datensätzen mit hoher Dichte.

In diesen Experimenten konnte das System neuronale Aktivität von 384 Aufzeichnungskanälen in Echtzeit erfolgreich digitalisieren, komprimieren, paketieren, serialisieren und rekonstruieren. Dank des „Send-on-Delta“-Ansatzes und des ternären, paketbasierten AER-Schemas von imec erzielte der Chip durchweg eine mehr als zehnfache Reduzierung des Datenvolumens, selbst unter Berücksichtigung des Overheads bei der Paketierung.

Entscheidend ist, dass dieser Kompressionsgrad ohne Einbußen bei der Spike-Genauigkeit erreicht wurde. Das System bewahrte Spikes mit Amplituden bis hinunter zu 31 μ V und rekonstruierte diese mit einem normalisierten RMS-Fehler von <23 % – was einem SNDR von 12,7 dB entspricht, was gut mit dem allgemein anerkannten Schwellenwert für zuverlässiges Spike-Sorting übereinstimmt. Mit anderen Worten: Der komprimierte, serialisierte Datenstrom behält alle Wellenformmerkmale bei, die für die nachgelagerte neuronale Dekodierung (und Analyse) unerlässlich sind.

NCT-Telemetrikette

Die gesamte NCT-Telemetrikette arbeitet mit außergewöhnlich geringem Stromverbrauch (nur 0,1 μ W pro Kanal) und weist eine

rekordverdächtige Siliziumeffizienz auf, da sie nur 27 Bit Speicher pro Kanal benötigt – eine 55-fache Reduzierung im Vergleich zu epochenbasierten Komprimierungsverfahren, die auf Kilobits an Pufferspeicher angewiesen sind. Dieser drastisch geringere Speicherbedarf minimiert die Siliziumfläche, senkt sowohl die Leckstrom- als auch die dynamische Leistungsaufnahme und trägt dazu bei, die Implantattemperaturen sicher innerhalb der klinischen Grenzwerte zu halten.

Deterministische Latenzen

Die Bedeutung deterministischer Latenzen in verteilten neuronalen Implantaten: Neuronale Spikes sind extrem kurz – oft deutlich kürzer als 200 μ s – und ihr genauer Zeitpunkt enthält wesentliche Informationen darüber, wie das Gehirn Bewegung, Wahrnehmung und Absichten codiert. In verteilten (intra-)kortikalen Systemen, in denen mehrere Aufzeichnungskanäle gleichzeitig Daten aus verschiedenen kortikalen Regionen erfassen, können selbst geringe Schwankungen in der Übertragungsverzögerung die zeitlichen Beziehungen zwischen den Spikes verzerren.

Um diese Beziehungen zu erhalten, muss das Telemetriesystem eine deterministische Latenz gewährleisten, wobei die zeitliche Ungenauigkeit auf nur wenige Mikrosekunden begrenzt bleibt.

NCT-Architektur

Die NCT-Architektur von Imec erfüllt diese Anforderung bereits durch ihre Konzeption. Durch die Beseitigung von Arbitrierungsverzögerungen und den Verzicht auf eine globale Taktverteilung stellt das System sicher, dass die Daten jeder Sensoreinheit in Echtzeit synchronisiert werden. Messungen zeigen eine Latenzschwankung von deutlich unter 10 μ s, womit die für die verteilte Spike-Timing-Analyse erforderliche Präzision im Mikrosekundenbereich problemlos erreicht wird.

Da die Anzahl der Aufzeichnungskanäle zunimmt und diese räumlich immer weiter verteilt sind, gewährleistet dieses deterministische Timing, dass die neuronale Aktivität über Tausende von Kanälen hinweg genau rekonstruiert werden kann, ohne zeitliche Abweichungen oder Verzerrungen.

Skalierung auf 10.000 Kanäle

Die jüngsten Ergebnisse von Imec zeigen, dass sich die neuromorphe kompressive Telemetrie-Architektur des Unternehmens bereits auf 1.500 Kanäle skalieren lässt – was den derzeitigen MEA-Plattformen mit der höchsten Kanaldichte entspricht –, während gleichzeitig eine 10-fache Datenreduktion erzielt und eine hochpräzise Rekonstruktion der Spike-Signale gewährleistet wird. Dies bestätigt, dass die Kernprinzipien – d. h. die ereignisgesteuerte „Send-on-Delta“-Signalerfassung sowie die ternäre AER-Paketierung/Linearisierung – weit über die anfänglichen Tests mit 384 Kanälen hinausreichen.

Der nächste Schritt

Als nächsten Schritt ergänzt das Team den NCT-Chip nun mit einem KI-gestützten Autoencoder, um die etwa 1 % der neuronalen Ereignisse zu identifizieren, die die größte verhaltensbezogene oder klinische Relevanz aufweisen. Durch die selektive Kodierung und Übertragung nur dieser besonders informativen Teilmenge soll die NCT-Architektur von imec eine 100-fache Datenreduktion erreichen – was eine praktische Skalierung auf 10.000 Aufzeichnungskanäle ermöglicht.

Imec lädt Entwickler von Systemen zur Erfassung elektrophysiologischer Daten, BCI-Unternehmen und Neurotech-Start-ups zur Zusammenarbeit ein – sei es durch die Integration des NCT-Chips von imec in ihre Plattformen oder durch die Lizenzierung der zugrunde liegenden IP-Blöcke für die (präklinische) Forschung. Durch die Zusammenarbeit lassen sich (i)BCIs der nächsten Generation schneller vorantreiben und skalierbare, klinisch einsetzbare neuronale Schnittstellen der Realität näherbringen.

Literatur:

Y. He et al., „An Event-Based Neural Compressive Telemetry With >11x Loss-Less Data Reduction for High-Bandwidth Intracortical Brain Computer Interfaces,“ in IEEE Transactions on Biomedical Circuits and Systems, vol. 18, no. 5, pp. 1100-1111, Oct. 2024, doi: 10.1109/TBCAS.2024.3378973.

Pietro Russo et al., „Event-based SerDes Telemetry Network for Distributed Brain Computer Interfaces,“ in 2026 Neuromorph. Comput. Eng. 6 014001, doi: 10.1088/2634-4386/ae2cc2

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 101001448). ◀